

Babel

Version 3.52.2228
2020/12/21

Original author
Johannes L. Braams

Current maintainer
Javier Bezos

Localization and
internationalization

Unicode

T_EX

pdfT_EX

LuaT_EX

XeT_EX

Contents

I	User guide	4
1	The user interface	4
1.1	Monolingual documents	4
1.2	Multilingual documents	6
1.3	Mostly monolingual documents	8
1.4	Modifiers	8
1.5	Troubleshooting	9
1.6	Plain	9
1.7	Basic language selectors	9
1.8	Auxiliary language selectors	10
1.9	More on selection	11
1.10	Shorthands	12
1.11	Package options	16
1.12	The base option	18
1.13	ini files	19
1.14	Selecting fonts	27
1.15	Modifying a language	29
1.16	Creating a language	30
1.17	Digits and counters	33
1.18	Dates	35
1.19	Accessing language info	35
1.20	Hyphenation and line breaking	37
1.21	Selection based on BCP 47 tags	39
1.22	Selecting scripts	40
1.23	Selecting directions	41
1.24	Language attributes	45
1.25	Hooks	46
1.26	Languages supported by babel with ldf files	47
1.27	Unicode character properties in luatex	48
1.28	Tweaking some features	49
1.29	Tips, workarounds, known issues and notes	49
1.30	Current and future work	50
1.31	Tentative and experimental code	50
2	Loading languages with language.dat	51
2.1	Format	51
3	The interface between the core of babel and the language definition files	52
3.1	Guidelines for contributed languages	53
3.2	Basic macros	54
3.3	Skeleton	55
3.4	Support for active characters	56
3.5	Support for saving macro definitions	56
3.6	Support for extending macros	57
3.7	Macros common to a number of languages	57
3.8	Encoding-dependent strings	57
4	Changes	61
4.1	Changes in babel version 3.9	61
II	Source code	61

5	Identification and loading of required files	62
6	locale directory	62
7	Tools	63
7.1	Multiple languages	67
7.2	The Package File (<code>\LaTeX</code> , <code>babel.sty</code>)	68
7.3	<code>base</code>	69
7.4	Conditional loading of shorthands	72
7.5	Cross referencing macros	73
7.6	Marks	76
7.7	Preventing clashes with other packages	77
7.7.1	<code>ifthen</code>	77
7.7.2	<code>varioref</code>	78
7.7.3	<code>hhline</code>	78
7.7.4	<code>hyperref</code>	79
7.7.5	<code>fancyhdr</code>	79
7.8	Encoding and fonts	79
7.9	Basic bidi support	81
7.10	Local Language Configuration	86
8	The kernel of Babel (<code>babel.def</code>, <code>common</code>)	91
8.1	Tools	91
9	Multiple languages	92
9.1	Selecting the language	94
9.2	Errors	103
9.3	Hooks	105
9.4	Setting up language files	107
9.5	Shorthands	110
9.6	Language attributes	119
9.7	Support for saving macro definitions	121
9.8	Short tags	122
9.9	Hyphens	122
9.10	Multiencoding strings	124
9.11	Macros common to a number of languages	131
9.12	Making glyphs available	131
9.12.1	Quotation marks	131
9.12.2	Letters	133
9.12.3	Shorthands for quotation marks	134
9.12.4	Umlauts and tremas	135
9.13	Layout	136
9.14	Load engine specific macros	137
9.15	Creating and modifying languages	137
10	Adjusting the Babel bahavior	157
11	Loading hyphenation patterns	158
12	Font handling with <code>fontspec</code>	163

13	Hooks for XeTeX and LuaTeX	168
13.1	XeTeX	168
13.2	Layout	170
13.3	LuaTeX	171
13.4	Southeast Asian scripts	177
13.5	CJK line breaking	181
13.6	Automatic fonts and ids switching	181
13.7	Layout	192
13.8	Auto bidi with basic and basic-r	195
14	Data for CJK	205
15	The ‘nil’ language	206
16	Support for Plain T_EX (plain.def)	206
16.1	Not renaming hyphen.tex	206
16.2	Emulating some L _A T _E X features	207
16.3	General tools	208
16.4	Encoding related macros	212
17	Acknowledgements	214

Troubleshooting

Paragraph ended before \UTFviii@three@octets was complete	5
No hyphenation patterns were preloaded for (babel) the language ‘LANG’ into the format	6
You are loading directly a language style	9
Unknown language ‘LANG’	9
Argument of \language@active@arg” has an extra }	13
Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’	29
Package babel Info: The following fonts are not babel standard families	29

Part I

User guide

What is this document about? This user guide focuses on internationalization and localization with \LaTeX and `pdftex`, `xetex` and `luatex` with the `babel` package. There are also some notes on its use with Plain \TeX . Part II describes the code, and usually it can be ignored.

What if I'm interested only in the latest changes? Changes and new features with relation to version 3.8 are highlighted with `New X.XX`, and there are some notes for the latest versions in [the babel wiki](#). The most recent features can be still unstable.

Can I help? Sure! If you are interested in the \TeX multilingual support, please join the [kadingira mail list](#). You can follow the development of `babel` in [GitHub](#) and make suggestions; feel free to fork it and make pull requests. If you are the author of a package, send to me a few test files which I'll add to mine, so that possible issues can be caught in the development phase.

It doesn't work for me! You can ask for help in some forums like [tex.stackexchange](#), but if you have found a bug, I strongly beg you to report it in [GitHub](#), which is much better than just complaining on an e-mail list or a web forum. Remember *warnings are not errors* by themselves, they just warn about possible problems or incompatibilities.

How can I contribute a new language? See section 3.1 for contributing a language.

I only need learn the most basic features. The first subsections (1.1-1.3) describe the traditional way of loading a language (with `ldf` files), which is usually all you need. The alternative way based on `ini` files, which complements the previous one (it does *not* replace it, although it is still necessary in some languages), is described below; go to 1.13.

I don't like manuals. I prefer sample files. This manual contains lots of examples and tips, but in [GitHub](#) there are many [sample files](#).

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. Another approach is making the language a global option in order to let other packages detect and use it. This is the standard way in \LaTeX for an option – in this case a language – to be recognized by several packages.

Many languages are compatible with `xetex` and `luatex`. With them you can use `babel` to localize the documents. When these engines are used, the Latin script is covered by default in current \LaTeX (provided the document encoding is UTF-8), because the font loader is preloaded and the font is switched to `lmroman`. Other scripts require loading `fontspec`. You may want to set the font attributes with `fontspec`, too.

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

PDFTEX

```
\documentclass{article}

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

\usepackage[french]{babel}

\begin{document}

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

\end{document}
```

Now consider something like:

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

With this setting, the package `varioref` will also see the option `french` and will be able to use it.

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

LUATEX/XETEX

```
\documentclass[russian]{article}

\usepackage{babel}

\babelfont{rm}{DejaVu Serif}

\begin{document}

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

\end{document}
```

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

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

Or the more explanatory:

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

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

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

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

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

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

NOTE With hyperref you may want to set the document language with something like:

```
\usepackage[pdflang=es-MX]{hyperref}
```

This is not currently done by babel and you must set it by hand.

NOTE Although it has been customary to recommend placing `\title`, `\author` and other elements printed by `\maketitle` after `\begin{document}`, mainly because of shorthands, it is advisable to keep them in the preamble. Currently there is no real need to use shorthands in those macros.

1.2 Multilingual documents

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

EXAMPLE In \LaTeX , the preamble of the document:

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

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

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

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

Examples of cases where main is useful are the following.

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

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

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

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

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

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

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

PDFTEX

```
\documentclass{article}  
  
\usepackage[T1]{fontenc}  
\usepackage[utf8]{inputenc}  
  
\usepackage[english,french]{babel}  
  
\begin{document}  
  
Plus ça change, plus c'est la même chose!  
  
\selectlanguage{english}  
  
And an English paragraph, with a short text in  
\foreignlanguage{french}{français}.  
  
\end{document}
```

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

LUATEX/XETEX

```
\documentclass{article}  
  
\usepackage[vietnamese,danish]{babel}  
  
\begin{document}  
  
\prefacename{} -- \alsoname{} -- \today
```



```

\selectlanguage{vietnamese}

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

\end{document}

```

NOTE Once loaded a language, you can select it with the corresponding BCP47 tag. See section 1.21 for further details.

1.3 Mostly monolingual documents

New 3.39 Very often, multilingual documents consist of a main language with small pieces of text in another languages (words, idioms, short sentences). Typically, all you need is to set the line breaking rules and, perhaps, the font. In such a case, babel now does not require declaring these secondary languages explicitly, because the basic settings are loaded on the fly when the language is selected (and also when provided in the optional argument of `\babelfont`, if used.)

This is particularly useful, too, when there are short texts of this kind coming from an external source whose contents are not known on beforehand (for example, titles in a bibliography). At this regard, it is worth remembering that `\babelfont` does *not* load any font until required, so that it can be used just in case.

EXAMPLE A trivial document is:

LUATEX/XETEX

```

\documentclass[english]{article}
\usepackage{babel}

\babelfont[russian]{rm}{FreeSerif}

\begin{document}

English. \foreignlanguage{russian}{Русский}.
\foreignlanguage{spanish}{Español}

\end{document}

```

NOTE Instead of its name, you may prefer to select the language with the corresponding BCP47 tag. This alternative, however, must be activated explicitly, because a two- or tree-letter word is a valid name for a language (eg, `yi`). See section 1.21 for further details.

1.4 Modifiers

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

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

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

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

1.5 Troubleshooting

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

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

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

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

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

1.6 Plain

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

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

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

1.7 Basic language selectors

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

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

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

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

```
\selectlanguage{german}
```

This command can be used as environment, too.

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

New 3.43 However, if the macro name does not match any language, it will get expanded as expected.

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

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

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

`\foreignlanguage` $[\langle option-list \rangle]\{\langle language \rangle\}\{\langle text \rangle\}$

The command `\foreignlanguage` takes two arguments; the second argument is a phrase to be typeset according to the rules of the language named in its first one.

This command (1) only switches the extra definitions and the hyphenation rules for the language, *not* the names and dates, (2) does not send information about the language to auxiliary files (i.e., the surrounding language is still in force), and (3) it works even if the language has not been set as package option (but in such a case it only sets the hyphenation patterns and a warning is shown). With the `bidi` option, it also enters in horizontal mode (this is not done always for backwards compatibility).

New 3.44 As already said, captions and dates are not switched. However, with the optional argument you can switch them, too. So, you can write:

```
\foreignlanguage[date]{polish}{\today}
```

In addition, captions can be switched with `captions` (or both, of course, with `date, captions`). Until 3.43 you had to write something like `{\selectlanguage{.} ...}`, which was not always the most convenient way.

1.8 Auxiliary language selectors

`\begin{otherlanguage}` $\{\langle language \rangle\} \dots \text{\code{\end{otherlanguage}}}$

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

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

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

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

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

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

`\begin{otherlanguage*}` [*<option-list>*] {<language>} ... `\end{otherlanguage*}`

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

This environment was originally intended for intermixing left-to-right typesetting with right-to-left typesetting in engines not supporting a change in the writing direction inside a line. However, by default it never complied with the documented behavior and it is just a version as environment of `\foreignlanguage`, except when the option `bidi` is set – in this case, `\foreignlanguage` emits a `\leavevmode`, while `other language*` does not.

`\begin{hyphenrules}` {<language>} ... `\end{hyphenrules}`

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

1.9 More on selection

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

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

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

EXAMPLE With

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

you can write

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

and

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

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

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

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

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

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

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

```
\babelensure{polish}
```

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

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

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

1.10 Shorthands

A *shorthand* is a sequence of one or two characters that expands to arbitrary \TeX code. Shorthands can be used for different kinds of things, 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.

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

The package `inputenc` as well as `xetex` and `luatex` have alleviated entering non-ASCII characters, but minority languages and some kinds of text can still require characters not directly available on the keyboards (and sometimes not even as separated or precomposed Unicode characters). As to the point 2, now pdfTeX provides `\knbcode`, and `luatex` can manipulate the glyph list. Tools for point 3 can be still very useful in general. There are three levels of shorthands: *user*, *language*, and *system* (by order of precedence). Version 3.9 introduces the *language user* level on top of the user level, as described below. In most cases, you will use only shorthands provided by languages.

NOTE Note the following:

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

TROUBLESHOOTING A typical error when using shorthands is the following:

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

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

`\shorthandon` `{\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.

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

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

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

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

`\usesshorthands` `*{\langle char \rangle}`

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

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

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

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

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

User shorthands override language ones, which in turn override system shorthands.

Language-dependent user shorthands (new in 3.9) take precedence over “normal” user shorthands.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Languages with only " as defined shorthand character Albanian, Bulgarian, Danish, Dutch, Finnish, German (old and new orthography, also Austrian), Icelandic, Italian, Norwegian, Polish, Portuguese (also Brazilian), Russian, Serbian (with Latin script), Slovene, Swedish, Ukrainian, Upper Sorbian

Basque " ' ~

Breton : ; ? !

Catalan " ' `

Czech " -

Esperanto ^

Estonian " ~

French (all varieties) : ; ? !

Galician " . ' ~ < >

Greek ~

Hungarian `

Kurmanji ^

Latin " ^ =

Slovak " ^ ' -

Spanish " . < > ' ~

Turkish : ! =

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

⁷Thanks to Enrico Gregorio

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

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

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

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

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

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

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

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

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

1.11 Package options

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

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

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

activegrave Same for ```.

shorthands= $\langle char \rangle \langle char \rangle \dots \mid \text{off}$

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

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

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

safe=	none ref bib
	Some \LaTeX macros are redefined so that using shorthands is safe. With <code>safe=bib</code> only <code>\nocite</code> , <code>\bibcite</code> and <code>\bibitem</code> are redefined. With <code>safe=ref</code> only <code>\newlabel</code> , <code>\ref</code> and <code>\pageref</code> are redefined (as well as a few macros from <code>varioref</code> and <code>ifthen</code>). With <code>safe=none</code> no macro is redefined. This option is strongly recommended, because a good deal of incompatibilities and errors are related to these redefinitions. As of New 3.34 , in $\epsilon\TeX$ based engines (ie, almost every engine except the oldest ones) shorthands can be used in these macros (formerly you could not).
math=	active normal
	Shorthands are mainly intended for text, not for math. By setting this option with the value <code>normal</code> they are deactivated in math mode (default is <code>active</code>) and things like <code>#{a'}</code> (a closing brace after a shorthand) are not a source of trouble anymore.
config=	$\langle file \rangle$
	Load $\langle file \rangle$.cfg instead of the default config file <code>bblopts.cfg</code> (the file is loaded even with <code>noconfigs</code>).
main=	$\langle language \rangle$
	Sets the main language, as explained above, ie, this language is always loaded last. If it is not given as package or global option, it is added to the list of requested languages.
headfoot=	$\langle language \rangle$
	By default, headlines and footlines are not touched (only marks), and if they contain language-dependent macros (which is not usual) there may be unexpected results. With this option you may set the language in heads and foots.
noconfigs	Global and language default config files are not loaded, so you can make sure your document is not spoilt by an unexpected .cfg file. However, if the key config is set, this file is loaded.
showlanguages	Prints to the log the list of languages loaded when the format was created: number (remember dialects can share it), name, hyphenation file and exceptions file.
nocase	New 3.9l Language settings for uppercase and lowercase mapping (as set by <code>\SetCase</code>) are ignored. Use only if there are incompatibilities with other packages.
silent	New 3.9l No warnings and no <i>infos</i> are written to the log file. ⁹
strings=	generic unicode encoded $\langle label \rangle$ $\langle font encoding \rangle$
	Selects the encoding of strings in languages supporting this feature. Predefined labels are <code>generic</code> (for traditional \TeX , LICR and ASCII strings), <code>unicode</code> (for engines like <code>xetex</code> and <code>luatex</code>) and <code>encoded</code> (for special cases requiring mixed encodings). Other allowed values are font encoding codes (T1, T2A, LGR, L7X...), but only in languages supporting them. Be aware with encoded captions are protected, but they work in <code>\MakeUpper</code> case and the like (this feature misuses some internal \LaTeX tools, so use it only as a last resort).
hyphenmap=	off first select other other*

⁹You can use alternatively the package `silence`.

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

off deactivates this feature and no case mapping is applied;
first sets it at the first switching commands in the current or parent scope (typically, when the aux file is first read and at `\begin{document}`), but also the first `\selectlanguage` in the preamble), and it's the default if a single language option has been stated;¹¹
select sets it only at `\selectlanguage`;
other also sets it at `otherlanguage`;
other* also sets it at `otherlanguage*` as well as in heads and foots (if the option `headfoot` is used) and in auxiliary files (ie, at `\select@language`), and it's the default if several language options have been stated. The option `first` can be regarded as an optimized version of `other*` for monolingual documents.¹²

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

New 3.14 Selects the bidi algorithm to be used in `luatex` and `xetex`. See sec. 1.23.

layout=

New 3.16 Selects which layout elements are adapted in bidi documents. See sec. 1.23.

1.12 The base option

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

\AfterBabelLanguage `{⟨option-name⟩}{⟨code⟩}`

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

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

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

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

¹⁰Turned off in plain.

¹¹Duplicated options count as several ones.

¹²Providing `foreign` is pointless, because the case mapping applied is that at the end of the paragraph, but if either `xetex` or `luatex` change this behavior it might be added. On the other hand, `other` is provided even if I [JBL] think it isn't really useful, but who knows.

```

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

```

WARNING Currently this option is not compatible with languages loaded on the fly.

1.13 ini files

An alternative approach to define a language (or, more precisely, a *locale*) is by means of an ini file. Currently babel provides about 200 of these files containing the basic data required for a locale.

ini files are not meant only for babel, and they have been devised as a resource for other packages. To ease interoperability between T_EX and other systems, they are identified with the BCP 47 codes as preferred by the Unicode Common Language Data Repository, which was used as source for most of the data provided by these files, too (the main exception being the `\ldf` name strings).

Most of them set the date, and many also the captions (Unicode and LICR). They will be evolving with the time to add more features (something to keep in mind if backward compatibility is important). The following section shows how to make use of them by means of `\babelprovide`. In other words, `\babelprovide` is mainly meant for auxiliary tasks, and as alternative when the `ldf`, for some reason, does not work as expected.

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

LUATEX/XETEX

```

\documentclass{book}

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

\babelfont{rm}[Renderer=Harfbuzz]{DejaVu Sans}

\begin{document}

\tableofcontents

\chapter{სამზარეულო და სუფრის ტრადიციები}

ქართული ტრადიციული სამზარეულო ერთ-ერთი უმდიდრესია მთელ მსოფლიოში.

\end{document}

```

New 3.49 Alternatively, you can tell babel to load all or some languages passed as options with `\babelprovide` and not from the `ldf` file in a few typical cases. Thus, `provide=*` means ‘load the main language with the `\babelprovide` mechanism instead of the `ldf` file’ applying the basic features, which in this case means `import, main`. There are (currently) three options:

- `provide=*` is the option just explained, for the main language;
- `provide+=*` is the same for additional languages (the main language is still the `ldf` file);
- `provide*=*` is the same for all languages, ie, main and additional.

EXAMPLE The preamble in the previous example can be more compactly written as:

```
\documentclass{book}
\usepackage[georgian, provide=*]{babel}
\babelfont{rm}[Renderer=Harfbuzz]{DejaVu Sans}
```

Or also:

```
\documentclass[georgian]{book}
\usepackage[provide=*]{babel}
\babelfont{rm}[Renderer=Harfbuzz]{DejaVu Sans}
```

NOTE The ini files just define and set some parameters, but the corresponding behavior is not always implemented. Also, there are some limitations in the engines. A few remarks follow (which could no longer be valid when you read this manual, if the packages involved have been updated). The Harfbuzz renderer has still some issues, so as a rule of thumb prefer the default renderer, and resort to Harfbuzz only if the former does not work for you. Fortunately, fonts can be loaded twice with different renderers; for example:

```
\babelfont[spanish]{rm}{FreeSerif}
\babelfont[hindi]{rm}[Renderer=Harfbuzz]{FreeSerif}
```

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

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

Devanagari In luatex and the the default renderer many fonts work, but some others do not, the main issue being the ‘ra’. You may need to set explicitly the script to either deva or dev2, eg:

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

Other Indic scripts are still under development in the default luatex renderer, but should work with Renderer=Harfbuzz. They also work with xetex, although fine tuning the font behavior is not always possible.

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

```
\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{lᦺ lᦴ lᦶ lᦸ lᦺ % Random
```

East Asia scripts Settings for either Simplified or Traditional should work out of the box, with basic line breaking with any renderer. Although for a few words and short texts the ini files should be fine, CJK texts are best set with a dedicated framework (CJK, luatexja, kotex, CTeX, etc.). This is what the class ltjbook does with luatex, which can be used in conjunction with the ldf for japanese, because the following piece of code loads luatexja:

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

Latin, Greek, Cyrillic Combining chars with the default luatex font renderer might be wrong; on the other hand, with the Harfbuzz renderer diacritics are stacked correctly, but many hyphenation points are discarded (this bug seems related to kerning, so it depends on the font). With xetex both combining characters and hyphenation work as expected (not quite, but in most cases it works; the problem here are font clusters).

NOTE Wikipedia defines a *locale* as follows: “In computing, a locale is a set of parameters that defines the user’s language, region and any special variant preferences that the user wants to see in their user interface. Usually a locale identifier consists of at least a language code and a country/region code.” Babel is moving gradually from the old and fuzzy concept of *language* to the more modern of *locale*. Note each locale is by itself a separate “language”, which explains why there are so many files. This is on purpose, so that possible variants can be created and/or redefined easily.

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

af	Afrikaans ^{ul}	cs	Czech ^{ul}
agq	Aghem	cu	Church Slavic
ak	Akan	cu-Cyrs	Church Slavic
am	Amharic ^{ul}	cu-Glag	Church Slavic
ar	Arabic ^{ul}	cy	Welsh ^{ul}
ar-DZ	Arabic ^{ul}	da	Danish ^{ul}
ar-MA	Arabic ^{ul}	dav	Taita
ar-SY	Arabic ^{ul}	de-AT	German ^{ul}
as	Assamese	de-CH	German ^{ul}
asa	Asu	de	German ^{ul}
ast	Asturian ^{ul}	dje	Zarma
az-Cyrl	Azerbaijani	dsb	Lower Sorbian ^{ul}
az-Latn	Azerbaijani	dua	Duala
az	Azerbaijani ^{ul}	dyo	Jola-Fonyi
bas	Basaa	dz	Dzongkha
be	Belarusian ^{ul}	ebu	Embu
bem	Bemba	ee	Ewe
bez	Bena	el	Greek ^{ul}
bg	Bulgarian ^{ul}	el-polyton	Polytonic Greek ^{ul}
bm	Bambara	en-AU	English ^{ul}
bn	Bangla ^{ul}	en-CA	English ^{ul}
bo	Tibetan ^u	en-GB	English ^{ul}
brx	Bodo	en-NZ	English ^{ul}
bs-Cyrl	Bosnian	en-US	English ^{ul}
bs-Latn	Bosnian ^{ul}	en	English ^{ul}
bs	Bosnian ^{ul}	eo	Esperanto ^{ul}
ca	Catalan ^{ul}	es-MX	Spanish ^{ul}
ce	Chechen	es	Spanish ^{ul}
cgg	Chiga	et	Estonian ^{ul}
chr	Cherokee	eu	Basque ^{ul}
ckb	Central Kurdish	ewo	Ewondo
cop	Coptic	fa	Persian ^{ul}

ff	Fulah	ksb	Shambala
fi	Finnish ^{ul}	ksf	Bafia
fil	Filipino	ksh	Colognian
fo	Faroese	kw	Cornish
fr	French ^{ul}	ky	Kyrgyz
fr-BE	French ^{ul}	lag	Langi
fr-CA	French ^{ul}	lb	Luxembourgish
fr-CH	French ^{ul}	lg	Ganda
fr-LU	French ^{ul}	lkt	Lakota
fur	Friulian ^{ul}	ln	Lingala
fy	Western Frisian	lo	Lao ^{ul}
ga	Irish ^{ul}	lrc	Northern Luri
gd	Scottish Gaelic ^{ul}	lt	Lithuanian ^{ul}
gl	Galician ^{ul}	lu	Luba-Katanga
grc	Ancient Greek ^{ul}	luo	Luo
gsw	Swiss German	luy	Luyia
gu	Gujarati	lv	Latvian ^{ul}
guz	Gusii	mas	Masai
gv	Manx	mer	Meru
ha-GH	Hausa	mfe	Morisyen
ha-NE	Hausa ¹	mg	Malagasy
ha	Hausa	mgf	Makhuwa-Meetto
haw	Hawaiian	mgo	Meta'
he	Hebrew ^{ul}	mk	Macedonian ^{ul}
hi	Hindi ^u	ml	Malayalam ^{ul}
hr	Croatian ^{ul}	mn	Mongolian
hsb	Upper Sorbian ^{ul}	mr	Marathi ^{ul}
hu	Hungarian ^{ul}	ms-BN	Malay ¹
hy	Armenian ^u	ms-SG	Malay ¹
ia	Interlingua ^{ul}	ms	Malay ^{ul}
id	Indonesian ^{ul}	mt	Maltese
ig	Igbo	mua	Mundang
ii	Sichuan Yi	my	Burmese
is	Icelandic ^{ul}	mzn	Mazanderani
it	Italian ^{ul}	naq	Nama
ja	Japanese	nb	Norwegian Bokmål ^{ul}
jgo	Ngomba	nd	North Ndebele
jmc	Machame	ne	Nepali
ka	Georgian ^{ul}	nl	Dutch ^{ul}
kab	Kabyle	nmg	Kwasio
kam	Kamba	nn	Norwegian Nynorsk ^{ul}
kde	Makonde	nnh	Ngiemboon
kea	Kabuverdianu	nus	Nuer
khq	Koyra Chiini	nyn	Nyankole
ki	Kikuyu	om	Oromo
kk	Kazakh	or	Odia
kkj	Kako	os	Ossetic
kl	Kalaallisut	pa-Arab	Punjabi
klj	Kalenjin	pa-Guru	Punjabi
km	Khmer	pa	Punjabi
kn	Kannada ^{ul}	pl	Polish ^{ul}
ko	Korean	pms	Piedmontese ^{ul}
kok	Konkani	ps	Pashto
ks	Kashmiri	pt-BR	Portuguese ^{ul}

pt-PT	Portuguese ^{ul}	sr	Serbian ^{ul}
pt	Portuguese ^{ul}	sv	Swedish ^{ul}
qu	Quechua	sw	Swahili
rm	Romansh ^{ul}	ta	Tamil ^u
rn	Rundi	te	Telugu ^{ul}
ro	Romanian ^{ul}	teo	Teso
rof	Rombo	th	Thai ^{ul}
ru	Russian ^{ul}	ti	Tigrinya
rw	Kinyarwanda	tk	Turkmen ^{ul}
rwk	Rwa	to	Tongan
sa-Beng	Sanskrit	tr	Turkish ^{ul}
sa-Deva	Sanskrit	twq	Tasawaq
sa-Gujr	Sanskrit	tzm	Central Atlas Tamazight
sa-Knda	Sanskrit	ug	Uyghur
sa-Mlym	Sanskrit	uk	Ukrainian ^{ul}
sa-Telu	Sanskrit	ur	Urdu ^{ul}
sa	Sanskrit	uz-Arab	Uzbek
sah	Sakha	uz-Cyrl	Uzbek
saq	Samburu	uz-Latn	Uzbek
sbp	Sangu	uz	Uzbek
se	Northern Sami ^{ul}	vai-Latn	Vai
seh	Sena	vai-Vaii	Vai
ses	Koyraboro Senni	vai	Vai
sg	Sango	vi	Vietnamese ^{ul}
shi-Latn	Tachelhit	vun	Vunjo
shi-Tfng	Tachelhit	wae	Walser
shi	Tachelhit	xog	Soga
si	Sinhala	yav	Yangben
sk	Slovak ^{ul}	yi	Yiddish
sl	Slovenian ^{ul}	yo	Yoruba
smn	Inari Sami	yue	Cantonese
sn	Shona	zgh	Standard Moroccan Tamazight
so	Somali		
sq	Albanian ^{ul}	zh-Hans-HK	Chinese
sr-Cyrl-BA	Serbian ^{ul}	zh-Hans-MO	Chinese
sr-Cyrl-ME	Serbian ^{ul}	zh-Hans-SG	Chinese
sr-Cyrl-XK	Serbian ^{ul}	zh-Hans	Chinese
sr-Cyrl	Serbian ^{ul}	zh-Hant-HK	Chinese
sr-Latn-BA	Serbian ^{ul}	zh-Hant-MO	Chinese
sr-Latn-ME	Serbian ^{ul}	zh-Hant	Chinese
sr-Latn-XK	Serbian ^{ul}	zh	Chinese
sr-Latn	Serbian ^{ul}	zu	Zulu

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

aghem	american
akan	amharic
albanian	ancientgreek

arabic	chinese-simplified-hongkongsarchina
arabic-algeria	chinese-simplified-macausarchina
arabic-DZ	chinese-simplified-singapore
arabic-morocco	chinese-simplified
arabic-MA	chinese-traditional-hongkongsarchina
arabic-syria	chinese-traditional-macausarchina
arabic-SY	chinese-traditional
armenian	chinese
assamese	churchslavic
asturian	churchslavic-cyrs
asu	churchslavic-oldcyrillic ¹³
australian	churchsslavic-glag
austrian	churchsslavic-glagolitic
azerbaijani-cyrillic	colognian
azerbaijani-cyrl	cornish
azerbaijani-latin	croatian
azerbaijani-latn	czech
azerbaijani	danish
bafia	duala
bambara	dutch
basaa	dzongkha
basque	embu
belarusian	english-au
bemba	english-australia
ben	english-ca
bengali	english-canada
bodo	english-gb
bosnian-cyrillic	english-newzealand
bosnian-cyrl	english-nz
bosnian-latin	english-unitedkingdom
bosnian-latn	english-unitedstates
bosnian	english-us
brazilian	english
breton	esperanto
british	estonian
bulgarian	ewe
burmese	ewondo
canadian	faroes
cantonese	filipino
catalan	finnish
centralatlastamazight	french-be
centralkurdish	french-belgium
chechen	french-ca
cherokee	french-canada
chiga	french-ch
chinese-hans-hk	french-lu
chinese-hans-mo	french-luxembourg
chinese-hans-sg	french-switzerland
chinese-hans	french
chinese-hant-hk	friulian
chinese-hant-mo	fulah
chinese-hant	galician

¹³The name in the CLDR is Old Church Slavonic Cyrillic, but it has been shortened for practical reasons.

ganda
georgian
german-at
german-austria
german-ch
german-switzerland
german
greek
gujarati
gusii
hausa-gh
hausa-ghana
hausa-ne
hausa-niger
hausa
hawaiian
hebrew
hindi
hungarian
icelandic
igbo
inarisami
indonesian
interlingua
irish
italian
japanese
jolafonyi
kabuverdianu
kabyle
kako
kalaallisut
kalenjin
kamba
kannada
kashmiri
kazakh
khmer
kikuyu
kinyarwanda
konkani
korean
koyraborosenni
koyrachiini
kwasio
kyrgyz
lakota
langi
lao
latvian
lingala
lithuanian
lowersorbian
lsorbian

lubakatanga
luo
luxembourgish
luyia
macedonian
machame
makhuwameetto
makonde
malagasy
malay-bn
malay-brunei
malay-sg
malay-singapore
malay
malayalam
maltese
manx
marathi
masai
mazanderani
meru
meta
mexican
mongolian
morisyen
mundang
nama
nepali
newzealand
ngiemboon
ngomba
norsk
northernluri
northernsami
northndebele
norwegianbokmal
norwegiannynorsk
nswissgerman
nuer
nyankole
nynorsk
occitan
oriya
oromo
ossetic
pashto
persian
piedmontese
polish
polytonicgreek
portuguese-br
portuguese-brazil
portuguese-portugal
portuguese-pt

portuguese	slovak
punjabi-arab	slovene
punjabi-arabic	slovenian
punjabi-gurmukhi	soga
punjabi-guru	somali
punjabi	spanish-mexico
quechua	spanish-mx
romanian	spanish
romansh	standardmoroccantamazight
rombo	swahili
rundi	swedish
russian	swissgerman
rwa	tachelhit-latin
sakha	tachelhit-latn
samburu	tachelhit-tfng
samin	tachelhit-tifinagh
sango	tachelhit
sangu	taita
sanskrit-beng	tamil
sanskrit-bengali	tasawaq
sanskrit-deva	telugu
sanskrit-devanagari	teso
sanskrit-gujarati	thai
sanskrit-gujr	tibetan
sanskrit-kannada	tigrinya
sanskrit-knda	tongan
sanskrit-malayalam	turkish
sanskrit-mlym	turkmen
sanskrit-telu	ukenglish
sanskrit-telugu	ukrainian
sanskrit	upporsorbian
scottishgaelic	urdu
sena	usenglish
serbian-cyrillic-bosniaherzegovina	usorbian
serbian-cyrillic-kosovo	uyghur
serbian-cyrillic-montenegro	uzbek-arab
serbian-cyrillic	uzbek-arabic
serbian-cyrl-ba	uzbek-cyrillic
serbian-cyrl-me	uzbek-cyrl
serbian-cyrl-xk	uzbek-latin
serbian-cyrl	uzbek-latn
serbian-latin-bosniaherzegovina	uzbek
serbian-latin-kosovo	vai-latin
serbian-latin-montenegro	vai-latn
serbian-latin	vai-vai
serbian-latn-ba	vai-vaii
serbian-latn-me	vai
serbian-latn-xk	vietnam
serbian-latn	vietnamese
serbian	vunjo
shambala	walser
shona	welsh
sichuanyi	westernfrisian
sinhala	yangben

yiddish
yoruba

zarma
zulu afrikaans

Modifying and adding values to ini files

New 3.39 There is a way to modify the values of ini files when they get loaded with `\babelprovide` and `import`. To set, say, `digits.native` in the `numbers` section, use something like `numbers/digits.native=abcdefghijkl`. Keys may be added, too. Without `import` you may modify the identification keys. This can be used to create private variants easily. All you need is to import the same ini file with a different locale name and different parameters.

1.14 Selecting fonts

New 3.15 Babel provides a high level interface on top of `fontspec` to select fonts. There is no need to load `fontspec` explicitly – babel does it for you with the first `\babelfont`.¹⁴

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

NOTE See the note in the previous section about some issues in specific languages.

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

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

If no language is given, then it is considered the default font for the family, activated when a language is selected.

On the other hand, if there is one or more languages in the optional argument, the font will be assigned to them, overriding the default one. Alternatively, you may set a font for a script – just precede its name (lowercase) with a star (eg, `*devanagari`). With this optional argument, the font is *not* yet defined, but just predeclared. This means you may define as many fonts as you want ‘just in case’, because if the language is never selected, the corresponding `\babelfont` declaration is just ignored.

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

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

LUATEX/XETEX

```
\documentclass{article}

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

\babelprovide[import]{hebrew}

\babelfont{rm}{FreeSerif}

\begin{document}
```

¹⁴See also the package `combofont` for a complementary approach.

```
Svenska \foreignlanguage{hebrew}{עברית} svenska.

\end{document}
```

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

LUATEX/XETEX

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

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

EXAMPLE Here is how to do it:

LUATEX/XETEX

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

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

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

LUATEX/XETEX

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

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

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

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

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

WARNING Using `\setxxxxfont` and `\babelfont` at the same time is discouraged, but very often works as expected. However, be aware with `\setxxxxfont` the language system will not be set by `babel` and should be set with `fontspec` if necessary.

TROUBLESHOOTING *Package fontspec Warning: ‘Language ‘LANG’ not available for font ‘FONT’ with script ‘SCRIPT’ ‘Default’ language used instead’.*

This is *not* and error. This warning is shown by fontspec, not by babel. It can be irrelevant for English, but not for many other languages, including Urdu and Turkish. This is a useful and harmless warning, and if everything is fine with your document the best thing you can do is just to ignore it altogether.

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

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

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

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

1.15 Modifying a language

Modifying the behavior of a language (say, the chapter “caption”), is sometimes necessary, but not always trivial. In the case of caption names a specific macro is provided, because this is perhaps the most frequent change:

`\setlocalecaption` $\{\langle\text{language-name}\rangle\}\{\langle\text{caption-name}\rangle\}\{\langle\text{string}\rangle\}$

New 3.51 Here *caption-name* is the name as string without the trailing name. An example, which also shows caption names are often a stylistic choice, is:

```
\setlocalecaption{english}{contents}{Table of Contents}
```

This works not only with existing caption names, because it also serves to define new ones by setting the *caption-name* to the name of your choice (name will be postpended). Captions so defined or redefined behave with the ‘new way’ described in the following note.

NOTE There are a few alternative methods:

- With data import’ed from ini files, you can modify the values of specific keys, like:

```
\babelprovide[import, captions/listtable = Lista de tablas]{spanish}
```

(In this particular case, instead of the captions group you may need to modify the `captions.licr` one.)

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

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

As of 3.15, there is no need to hide spaces with % (babel removes them), but it is advisable to do so. This redefinition is not activated until the language is selected.

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

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

This redefinition is immediate.

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

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

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

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

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

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

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

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

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

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

1.16 Creating a language

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

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

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

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

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

```

Package babel Warning: \mylangchaptername not set. Please, define it
(babel)                after the language has been loaded (typically
(babel)                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. Note languages loaded on the fly are not yet available in the preamble.

EXAMPLE If you need a language named arhinish:

```

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

```

EXAMPLE Locales with names based on BCP 47 codes can be created with something like:

```

\babelprovide[import=en-US]{enUS}

```

Note, however, mixing ways to identify locales can lead to problems. For example, is yi the name of the language spoken by the Yi people or is it the code for Yiddish?

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

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

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

import= *<language-tag>*

New 3.13 Imports data from an ini file, including captions and date (also line breaking rules in newly defined languages). For example:

```

\babelprovide[import=hu]{hungarian}

```

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

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

```

\babelprovide[import]{hungarian}

```

There are about 250 ini files, with data taken from the ldf files and the CLDR provided by Unicode. Not all languages in the latter are complete, and therefore neither are the ini files. A few languages may show a warning about the current lack of suitability of some features.

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

`\<language>date{\the\year}{\the\month}{\the\day}`. **New 3.44** More convenient is usually `\localedate`, which prints the date for the current locale.

captions= $\langle\textit{language-tag}\rangle$

Loads only the strings. For example:

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

hyphenrules= $\langle\textit{language-list}\rangle$

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

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

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

A special value is +, which allocates a new language (in the $\text{T}_{\text{E}}\text{X}$ 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 (thus overriding that set when babel is loaded). Only in newly defined languages.

EXAMPLE Let's assume your document is mainly in Polytonic Greek, but with some sections in Italian. Then, the first attempt should be:

```
\usepackage[italian, greek.polutonic]{babel}
```

But if, say, accents in Greek are not shown correctly, you can try:

```
\usepackage[italian]{babel}  
\babelprovide[import, main]{polytonicgreek}
```

script= $\langle\textit{script-name}\rangle$

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

language= $\langle\textit{language-name}\rangle$

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

alph= $\langle counter-name \rangle$

Assigns to `\alph` that counter. See the next section.

Alph= $\langle counter-name \rangle$

Same for `\Alph`.

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

onchar= `ids | fonts`

New 3.38 This option is much like an ‘event’ called when a character belonging to the script of this locale is found (as its name implies, it acts on characters, not on spaces). There are currently two ‘actions’, which can be used at the same time (separated by a space): with `ids` the `\language` and the `\localeid` are set to the values of this locale; with `fonts`, the fonts are changed to those of this locale (as set with `\babelfont`). This option is not compatible with `mapfont`. Characters can be added or modified with `\babelcharproperty`.

NOTE An alternative approach with `luatex` and `Harfbuzz` is the font option `RawFeature={multiscript=auto}`. It does not switch the `babel` language and therefore the line breaking rules, but in many cases it can be enough.

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

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

intrapenalty= $\langle penalty \rangle$

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

mapfont= `direction`

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

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

1.17 Digits and counters

New 3.20 About thirty `ini` files define a field named `digits.native`. When it is present, two macros are created: `\<language>digits` and `\<language>counter` (only `xetex` and

luatex). With the first, a string of ‘Latin’ digits are converted to the native digits of that language; the second takes a counter name as argument. With the option `maparabic` in `\babelprovide`, `\arabic` is redefined to produce the native digits (this is done *globally*, to avoid inconsistencies in, for example, page numbering, and note as well dates do not rely on `\arabic`.)

For example:

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

Languages providing native digits in all or some variants are:

Arabic	Persian	Lao	Odia	Urdu
Assamese	Gujarati	Northern Luri	Punjabi	Uzbek
Bangla	Hindi	Malayalam	Pashto	Vai
Tibetar	Khmer	Marathi	Tamil	Cantonese
Bodo	Kannada	Burmese	Telugu	Chinese
Central Kurdish	Konkani	Mazanderani	Thai	
Dzongkha	Kashmiri	Nepali	Uyghur	

New 3.30 With `luatex` there is an alternative approach for mapping digits, namely, `mapdigits`. Conversion is based on the language and it is applied to the typeset text (not math, PDF bookmarks, etc.) before bidi and fonts are processed (ie, to the node list as generated by the \TeX code). This means the local digits have the correct bidirectional behavior (unlike `Numbers=Arabic` in `fontspec`, which is not recommended).

NOTE With `xetex` you can use the option `Mapping` when defining a font.

New 4.41 Many ‘ini’ locale files has been extended with information about non-positional numerical systems, based on those predefined in CSS. They only work with `xetex` and `luatex` and are fully expendable (even inside an unprotected `\edef`). Currently, they are limited to numbers below 10000.

There are several ways to use them (for the available styles in each language, see the list below):

- `\localnumeral{<style>}{<number>}`, like `\localnumeral{abjad}{15}`
- `\localecounter{<style>}{<counter>}`, like `\localecounter{lower}{section}`
- In `\babelprovide`, as an argument to the keys `alph` and `Alph`, which redefine what `\alph` and `\Alph` print. For example:

```
\babelprovide[alph=alphabetic]{thai}
```

The styles are:

Ancient Greek `lower.ancient`, `upper.ancient`

Amharic `afar`, `agaw`, `ari`, `blin`, `dizi`, `gedeo`, `gumuz`, `hadiyya`, `harari`, `kaffa`, `kebena`, `kembata`, `konso`, `kunama`, `meen`, `oromo`, `saho`, `sidama`, `silti`, `tigre`, `wolaita`, `yemsa`

Arabic abjad, maghrebi.abjad
Belarusan, Bulgarian, Macedonian, Serbian lower, upper
Bengali alphabetic
Coptic epact, lower.letters
Hebrew letters (neither geresh nor gershayim yet)
Hindi alphabetic
Armenian lower.letter, upper.letter
Japanese hiragana, hiragana.iroha, katakana, katakana.iroha, circled.katakana, informal, formal, cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha
Georgian letters
Greek lower.modern, upper.modern, lower.ancient, upper.ancient (all with keraia)
Khmer consonant
Korean consonant, syllable, hanja.informal, hanja.formal, hangul.formal, cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha
Marathi alphabetic
Persian abjad, alphabetic
Russian lower, lower.full, upper, upper.full
Syriac letters
Tamil ancient
Thai alphabetic
Ukrainian lower, lower.full, upper, upper.full
Chinese cjk-earthly-branch, cjk-heavenly-stem, fullwidth.lower.alpha, fullwidth.upper.alpha

New 3.45 In addition, native digits (in languages defining them) may be printed with the numeral style digits.

1.18 Dates

New 3.45 When the data is taken from an ini file, you may print the date corresponding to the Gregorian calendar and other lunisolar systems with the following command.

`\localedate` [`<calendar=.., variant=..>`]{`<year>`}{`<month>`}{`<day>`}

By default the calendar is the Gregorian, but a ini files may define strings for other calendars (currently ar, ar-*, he, fa, hi.) In the latter case, the three arguments are the year, the month, and the day in those in the corresponding calendar. They are *not* the Gregorian data to be converted (which means, say, 13 is a valid month number with calendar=hebrew).

Even with a certain calendar there may be variants. In Kurmanji the default variant prints something like *30. Çileyê Pêşîn 2019*, but with variant=iza fa it prints *31'ê Çileyê Pêşînê 2019*.

1.19 Accessing language info

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

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

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

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

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

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

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

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

`tag.bcp47` is the full BCP 47 tag (see the warning below).

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

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

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

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

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

WARNING **New 3.46** As of version 3.46 `tag.bcp47` returns the full BCP 47 tag. Formerly it returned just the language subtag, which was clearly counterintuitive.

`\getlocaleproperty` $*\{\langle macro \rangle\}\{\langle locale \rangle\}\{\langle property \rangle\}$

New 3.42 The value of any locale property as set by the ini files (or added/modified with `\babelprovide`) can be retrieved and stored in a macro with this command. For example, after:

```
\getlocaleproperty\hechap{hebrew}{captions/chapter}
```

the macro `\hechap` will contain the string פרק.

If the key does not exist, the macro is set to `\relax` and an error is raised. **New 3.47** With the starred version no error is raised, so that you can take your own actions with undefined properties.

Babel remembers which ini files have been loaded. There is a loop named

`\LocaleForEach` to traverse the list, where #1 is the name of the current item, so that

`\LocaleForEach{\message{ **#1** }} just shows the loaded ini's.`

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

`\localeid`

Each language in the babel sense has its own unique numeric identifier, which can be retrieved with `\localeid`.

NOTE The `\localeid` is not the same as the `\language` identifier, which refers to a set of hyphenation patterns (which, in turn, is just a component of the line breaking algorithm described in the next section). The data about preloaded patterns are stored in an internal macro named `\bbl@languages` (see the code for further details), but note several locales may share a single `\language`, so they are separated concepts. In `luatex`, the `\localeid` is saved in each node (where it makes sense) as an attribute, too.

1.20 Hyphenation and line breaking

Babel deals with three kinds of line breaking rules: Western, typically the LGC group, South East Asian, like Thai, and CJK, but support depends on the engine: pdfTeX only deals with the former, xetex also with the second one (although in a limited way), while luatex provides basic rules for the latter, too.

`\babelhyphen` `*{<type>}`
`\babelhyphen` `*{<text>}`

New 3.9a It is customary to classify hyphens in two types: (1) *explicit* or *hard hyphens*, which in T_EX 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 T_EX 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 T_EX, - and \- forbid further breaking opportunities in the word. This is the desired behavior very often, but not always, and therefore many languages provide shorthands for these cases. Unfortunately, this has not been done consistently: for example, - in Dutch, Portuguese, Catalan or Danish is a hard hyphen, while in German, Spanish, Norwegian, Slovak or Russian is a soft hyphen. Furthermore, some of them even redefine \-, so that you cannot insert a soft hyphen without breaking opportunities in the rest of the word. Therefore, some macros are provided with a set of basic “hyphens” which can be used by themselves, to define a user shorthand, or even in language files.

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

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

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

There are also some differences with 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 `\babelnullhyphen`; (3) a break after the hyphen is forbidden if preceded by a glue >0 pt (at the beginning of a word, provided it is not immediately preceded by, say, a parenthesis).

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

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

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

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

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

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

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

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

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

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

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

\babelposthyphenation {<hyphenrules-name>} {<lua-pattern>} {<replacement>}

New 3.37-3.39 *With luatex* it is now possible to define non-standard hyphenation rules, like `f-f → ff-f`, repeated hyphens, ranked ruled (or more precisely, ‘penalized’ hyphenation points), and so on. No rules are currently provided by default, but they can be defined as shown in the following example, where `{1}` is the first captured char (between `()` in the pattern):

```
\babelposthyphenation{german}{([fmtrp]) | {1}}
{
  { no = {1}, pre = {1}{1}- }, % Replace first char with disc
  remove,                    % Remove automatic disc (2nd node)
  {}                          % Keep last char, untouched
}
```

In the replacements, a captured char may be mapped to another, too. For example, if the first capture reads `([îú])`, the replacement could be `{1|îú|íú}`, which maps `î` to `í`, and `ú` to `ó`, so that the diaeresis is removed.

This feature is activated with the first `\babelposthyphenation` or `\babelprehyphenation`.

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.

See the [babel wiki](#) for a more detailed description and some examples. It also describes a few additional replacement types (string, penalty).

Although the main purpose of this command is non-standard hyphenation, it may actually be used for other transformations (after hyphenation is applied, so you must take discretionaries into account).

You are limited to substitutions as done by lua, although a future implementation may alternatively accept lpeg.

`\babelprehyphenation` `{⟨locale-name⟩}{⟨lua-pattern⟩}{⟨replacement⟩}`

New 3.44-3-52 This command is not strictly about hyphenation, but it is include here because it is a clear counterpart of `\babelposthyphenation`. It is similar to the latter, but (as its name implies) applied before hyphenation. There are other differences: (1) the first argument is the locale instead the name of hyphenation patterns; (2) in the search patterns = has no special meaning, while | stands for an ordinary space; (3) in the replacement, discretionaries are not accepted.

It handles glyphs and spaces (but you can not insert spaces).

Performance is still somewhat poor in some cases, but it is fast in the most the typical ones. This feature is activated with the first `\babelposthyphenation` or `\babelprehyphenation`.

EXAMPLE You can replace a character (or series of them) by another character (or series of them). Thus, to enter ž as zh and š as sh in a newly created locale for transliterated Russian:

```
\babelprovide[hyphenrules=+]{russian-latin} % Create locale
\babelprehyphenation{russian-latin}{([sz])h} % Create rule
{
  string = {1|sz|šž},
  remove
}
```

EXAMPLE The following rule prevent the word “a” from being at the end of a line:

```
\babelprehyphenation{english}{|a|}
{ }, { }, % Keep first space and a
{insert, penalty = 10000}, % Insert penalty
{ } % Keep last space
}
```

1.21 Selection based on BCP 47 tags

New 3.43 The recommended way to select languages is that described at the beginning of this document. However, BCP 47 tags are becoming customary, particularly in documents (or parts of documents) generated by external sources, and therefore babel will provide a set of tools to select the locales in different situations, adapted to the particular needs of each case. Currently, babel provides autoloading of locales as described in this section. In these contexts autoloading is particularly important because we may not know on beforehand which languages will be requested.

It must be activated explicitly, because it is primarily meant for special tasks. Mapping from BCP 47 codes to locale names are not hardcoded in babel. Instead the data is taken from the ini files, which means currently about 250 tags are already recognized. Babel performs a simple lookup in the following way: fr-Latn-FR → fr-Latn → fr-FR → fr. Languages with the same resolved name are considered the same. Case is normalized

before, so that `fr-latn-fr` \rightarrow `fr-Latn-FR`. If a tag and a name overlap, the tag takes precedence.

Here is a minimal example:

```
\documentclass{article}

\usepackage[danish]{babel}

\babeladjust{
  autoload.bcp47 = on,
  autoload.bcp47.options = import
}

\begin{document}

Chapter in Danish: \chaptername.

\selectlanguage{de-AT}

\localedate{2020}{1}{30}

\end{document}
```

Currently the locales loaded are based on the `ini` files and decoupled from the main `ldf` files. This is by design, to ensure code generated externally produces the same result regardless of the languages requested in the document, but an option to use the `ldf` instead will be added in a future release, because both options make sense depending on the particular needs of each document (there will be some restrictions, however). The behaviour is adjusted with `\babeladjust` with the following parameters:

`autoload.bcp47` with values `on` and `off`.

`autoload.bcp47.options`, which are passed to `\babelprovide`; empty by default, but you may add `import` (features defined in the corresponding `babel-...tex` file might not be available).

`autoload.bcp47.prefix`. Although the public name used in selectors is the tag, the internal name will be different and generated by prepending a prefix, which by default is `bcp47-`. You may change it with this key.

New 3.46 If an `ldf` file has been loaded, you can enable the corresponding language tags as selector names with:

```
\babeladjust{ bcp47.toname = on }
```

(You can deactivate it with `off`.) So, if `dutch` is one of the package (or class) options, you can write `\selectlanguage{nl}`. Note the language name does not change (in this example is still `dutch`), but you can get it with `\localeinfo` or `\getlanguageproperty`. It must be turned on explicitly for similar reasons to those explained above.

1.22 Selecting scripts

Currently `babel` provides no standard interface to select scripts, because they are best selected with either `\fontencoding` (low-level) or a language name (high-level). Even the

Latin script may require different encodings (ie, sets of glyphs) depending on the language, and therefore such a switch would be in a sense incomplete.¹⁶

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

`\ensureascii` $\langle text \rangle$

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

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

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

1.23 Selecting directions

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

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

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

WARNING If characters to be mirrored are shown without changes with `luatex`, try with the following line:

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

¹⁷But still defined for backwards compatibility.

```
\babeladjust{bidi.mirroring=off}
```

There are some package options controlling bidi writing.

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

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

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

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

There are samples on GitHub, under /required/babel/samples. See particularly lua-bidibasic.tex and lua-secenum.tex.

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

```
\documentclass{article}

\usepackage[bidi=basic]{babel}

\babelprovide[import, main]{arabic}

\babelfont{rm}{FreeSerif}

\begin{document}

    وقد عرفت شبه جزيرة العرب طيلة العصر الهيليني (الاجريقي) بـ
    Arabia أو Aravia (بالاغريقية Αραβία)، استخدم الرومان ثلاث
    بادئات بـ "Arabia" على ثلاث مناطق من شبه الجزيرة العربية، إلا أنها
    حقيقةً كانت أكبر مما تعرف عليه اليوم.

\end{document}
```

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

```
\documentclass{book}

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

\babelprovide[onchar=ids fonts]{arabic}
```

```

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

\begin{document}

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

\end{document}

```

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

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

```

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

```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

graphics modifies the `picture` environment so that the whole figure is L but the text is R. It *does not* work with the standard `picture`, and `pict2e` is required if you want sloped lines (**With recent versions of L^AT_EX, this feature has stopped working**). It attempts to do the same for `pgf/tikz`. Somewhat experimental. **New 3.32** .

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

EXAMPLE Typically, in an Arabic document you would need:

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

\babelsublr `{\lr-text}`

Digits in `pdftex` must be marked up explicitly (unlike `luatex` with `bidi=basic` or `bidi=basic-r` and, usually, `xetex`). This command is provided to set `{\lr-text}` in L mode if necessary. It's intended for what Unicode calls weak characters, because words are best set with the corresponding language. For this reason, there is no `rl` counterpart. Any `\babelsublr` in *explicit* L mode is ignored. However, with `bidi=basic` and *implicit* L, it first returns to R and then switches to explicit L. To clarify this point, consider, in an R context:

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

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

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

\BabelPatchSection `{\langle section-name \rangle}`

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

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

New 3.17 Something like:

```
\BabelFootnote{\parsfootnote}{\language name}{\{ \}}
```

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

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

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

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

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

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

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

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

1.24 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).

1.25 Hooks

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

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

The same name can be applied to several events. Hooks may be enabled and disabled for all defined events with `\EnableBabelHook{<name>}`, `\DisableBabelHook{<name>}`. Names containing the string `babel` are reserved (they are used, for example, by `\useshortands*` to add a hook for the event `afterextras`). **New 3.33** They may be also applied to a specific language with the optional argument; language-specific settings are executed after global ones.

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

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

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

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

defaultcommands Used (locally) in `\StartBabelCommands`.

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

stopcommands Used to reset the above, if necessary.

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

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

afterextras Just after executing `\extras<language>`. For example, the following deactivates 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 just defines a few basic commands. It can be used to define different versions of them or to load a file.

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

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

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

1.26 Languages supported by babel with ldf files

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

Afrikaans afrikaans

Azerbaijani azerbaijani

Basque basque

Breton breton

Bulgarian bulgarian

Catalan catalan

Croatian croatian

Czech czech

Danish danish

Dutch dutch

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

Esperanto esperanto

Estonian estonian

Finnish finnish

French french, francais, canadien, acadian

Galician galician

German austrian, german, germanb, ngerman, naustrian

Greek greek, polutonikogreek

Hebrew hebrew

Icelandic icelandic

Indonesian indonesian (bahasa, indon, bahasai)

Interlingua interlingua

Irish Gaelic irish

Italian italian

Latin latin

Lower Sorbian lowersorbian

Malay malay, melayu (bahasam)

North Sami samin

Norwegian norsk, nynorsk

Polish polish

Portuguese portuguese, brazilian (portuges, brazil)¹⁹

¹⁹The two last name comes from the times when they had to be shortened to 8 characters

Romanian romanian
Russian russian
Scottish Gaelic scottish
Spanish spanish
Slovakian slovak
Slovenian slovene
Swedish swedish
Serbian serbian
Turkish turkish
Ukrainian ukrainian
Upper Sorbian uppsorbian
Welsh welsh

There are more languages not listed above, including hindi, thai, thaicjk, latvian, turkmen, magyar, mongolian, romansh, lithuanian, spanglish, vietnamese, japanese, pinyin, arabic, farsi, ibygreek, bgreek, serbianc, frenchle, ethiop and friulan. Most of them work out of the box, but some may require extra fonts, encoding files, a preprocessor or even a complete framework (like CJK or luatexja). For example, if you have got the velthuis/devnag package, you can create a file with extension .dn:

```

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

```

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

1.27 Unicode character properties in luatex

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

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

New 3.32 Here, $\{\langle char-code \rangle\}$ is a number (with \TeX syntax). With the optional argument, you can set a range of values. There are three properties (with a short name, taken from Unicode): direction (bc), mirror (bmg), linebreak (lb). The settings are global, and this command is allowed only in vertical mode (the preamble or between paragraphs). For example:

```

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

```

New 3.39 Another property is locale, which adds characters to the list used by onchar in $\backslash\text{babelprovide}$, or, if the last argument is empty, removes them. The last argument is the locale name:

```
\babelcharproperty{`,`}{locale}{english}
```

1.28 Tweaking some features

`\babeladjust` $\langle\text{key-value-list}\rangle$

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

1.29 Tips, workarounds, known issues and notes

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

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

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

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

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

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

- For the hyphenation to work correctly, `lccodes` cannot change, because \TeX only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.²⁰ So, if you write a chunk of French text with `\foreignlanguage`, the apostrophes might not be taken into account. This is a limitation of \TeX , not of `babel`. Alternatively, you may use `\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.

²⁰This explains why \LaTeX 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.

- Babel does not take into account `\normalsfcodes` and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the ‘to do’ list).
- Using a character mathematically active (ie, with math code "8000) as a shorthand can make \TeX enter in an infinite loop in some rare cases. (Another issue in the ‘to do’ list, although there is a partial solution.)

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

csquotes Logical markup for quotes.
iflang Tests correctly the current language.
hyphsubst Selects a different set of patterns for a language.
translator An open platform for packages that need to be localized.
siunitx Typesetting of numbers and physical quantities.
biblatex Programmable bibliographies and citations.
bicaption Bilingual captions.
babelbib Multilingual bibliographies.
microtype Adjusts the typesetting according to some languages (kerning and spacing). Ligatures can be disabled.
substitutefont Combines fonts in several encodings.
mkpattern Generates hyphenation patterns.
tracklang Tracks which languages have been requested.
ucharclasses (xetex) Switches fonts when you switch from one Unicode block to another.
zhspacing Spacing for CJK documents in xetex.

1.30 Current and future work

The current work is focused on the so-called complex scripts in luatex . In 8-bit engines, babel provided a basic support for bidi text as part of the style for Hebrew, but it is somewhat unsatisfactory and internally replaces some hardwired commands by other hardwired commands (generic changes would be much better). Useful additions would be, for example, time, currency, addresses and personal names.²¹ But that is the easy part, because they don’t require modifying the \TeX internals. Calendars (Arabic, Persian, Indic, etc.) are under study. Also interesting are differences in the sentence structure or related to it. For example, in Basque the number precedes the name (including chapters), in Hungarian “from (1)” is “(1)-ből”, but “from (3)” is “(3)-ból”, in Spanish an item labelled “3.^o” may be referred to as either “ítem 3.^o” or “3.^{er} ítem”, and so on. An option to manage bidirectional document layout in luatex (lists, footnotes, etc.) is almost finished, but xetex required more work. Unfortunately, proper support for xetex requires patching somehow lots of macros and packages (and some issues related to `\specials` remain, like color and hyperlinks), so babel resorts to the bidi package (by Vafa Khalighi). See the babel repository for a small example (xe-bidi).

1.31 Tentative and experimental code

See the code section for `\foreignlanguage*` (a new starred version of `\foreignlanguage`). For old an deprecated functions, see the wiki.

Options for locales loaded on the fly

New 3.51 `\babeladjust{ autoload.options = ... }` sets the options when a language is loaded on the fly (by default, no options). A typical value would be `import`, which

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

defines captions, date, numerals, etc., but ignores the code in the tex file (for example, extended numerals in Greek).

Labels

New 3.48 There is some work in progress for babel to deal with labels, both with the relation to captions (chapters, part), and how counters are used to define them. It is still somewhat tentative because it is far from trivial – see the wiki for further details.

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).²² Until 3.9n, this task was delegated to the package luatex-hyphen, by Khaled Hosny, Élie Roux, and Manuel Pégourié-Gonnard, and required an extra file named language.dat.lua, but now a new mechanism has been devised based solely on language.dat. **You must rebuild the formats** if upgrading from a previous version. You may want to have a local language.dat for a particular project (for example, a book on Chemistry).²³

2.1 Format

In that file the person who maintains a T_EX environment has to record for which languages he has hyphenation patterns *and* in which files these are stored²⁴. When hyphenation exceptions are stored in a separate file this can be indicated by naming that file *after* the file with the hyphenation patterns.

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

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

You may also set the font encoding the patterns are intended for by following the language name by a colon and the encoding code.²⁵ For example:

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

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

²³The loader for lua(e)tex is slightly different as it's not based on babel but on etex.src. Until 3.9p it just didn't work, but thanks to the new code it works by reloading the data in the babel way, i.e., with language.dat.

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

²⁵This is not a new feature, but in former versions it didn't work correctly.

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

Some recommendations:

- The preferred shorthand is `"`, which is not used in \LaTeX (quotes are entered as `` `` and `' '`). Other good choices are characters which are not used in a certain context (eg, `=` in an ancient language). Note however `=`, `<`, `>`, `:` and the like can be dangerous, because they may be used as part of the syntax of some elements (numeric expressions, key/value pairs, etc.).
- Captions should not contain shorthands or encoding-dependent commands (the latter is not always possible, but should be clearly documented). They should be defined using the LICR. You may also use the new tools for encoded strings, described below.
- Avoid adding things to `\noextras<lang>` except for `umlauthigh` and `friends`, `\bbl@deactivate`, `\bbl@(non)frenchspacing`, and language-specific macros. Use always, if possible, `\bbl@save` and `\bbl@savevariable` (except if you still want to have access to the previous value). Do not reset a macro or a setting to a hardcoded value. Never. Instead save its value in `\extras<lang>`.
- Do not switch scripts. If you want to make sure a set of glyphs is used, switch either the font encoding (low-level) or the language (high-level, which in turn may switch the font encoding). Usage of things like `\latintext` is deprecated.²⁶
- Please, for “private” internal macros do not use the `\bbl@` prefix. It is used by `babel` and it can lead to incompatibilities.

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

3.1 Guidelines for contributed languages

Currently, the easiest way to contribute a new language is by taking one of the 500 or so `ini` templates available on GitHub as a basis. Just make a pull request or download it and then, after filling the fields, send it to me. Feel free to ask for help or to make feature requests.

As to `ldf` files, now language files are “outsourced” and are located in a separate directory (`/macros/latex/contrib/babel-contrib`), so that they are contributed directly to CTAN (please, do not send to me language styles just to upload them to CTAN).

Of course, placing your style files in this directory is not mandatory, but if you want to do it, here are a few guidelines.

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

²⁶But not removed, for backward compatibility.

The following page provides a starting point for ldf files:

<http://www.texnia.com/incubator.html>. See also

<https://github.com/latex3/babel/wiki/List-of-locale-templates>.

If you need further assistance and technical advice in the development of language styles, I am willing to help you. And of course, you can make any suggestion you like.

3.2 Basic macros

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

\addlanguage The macro `\addlanguage` is a non-outer version of the macro `\newlanguage`, defined in `plain.tex` version 3.x. Here “language” is used in the T_EX sense of set of hyphenation patterns.

\adddialect The macro `\adddialect` can be used when two languages can (or must) use the same hyphenation patterns. This can also be useful for languages for which no patterns are preloaded in the format. In such cases the default behavior of the babel system is to define 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.

<code>\ldf@quit</code>	The macro <code>\ldf@quit</code> does work needed if a <code>.ldf</code> file was processed earlier. This includes resetting the category code of the <code>@</code> -sign, preparing the language to be activated at <code>\begin{document}</code> time, and ending the input stream.
<code>\ldf@finish</code>	The macro <code>\ldf@finish</code> does work needed at the end of each <code>.ldf</code> file. This includes resetting the category code of the <code>@</code> -sign, loading a local configuration file, and preparing the language to be activated at <code>\begin{document}</code> time.
<code>\loadlocalcfg</code>	After processing a language definition file, \TeX can be instructed to load a local configuration file. This file can, for instance, be used to add strings to <code>\captions{lang}</code> to support local document classes. The user will be informed that this configuration file has been loaded. This macro is called by <code>\ldf@finish</code> .
<code>\substitutefontfamily</code>	(Deprecated.) This command takes three arguments, a font encoding and two font family names. It creates a font description file for the first font in the given encoding. This <code>.fd</code> file will instruct \TeX to use a font from the second family when a font from the first family in the given encoding seems to be needed.

3.3 Skeleton

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

```

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

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

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

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

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

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

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

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

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

\EndBabelCommands

```



```

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

\ldf@finish{<language>}

```

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

```

\AtEndOfPackage{%
  \RequirePackage{dingbat}%      Delay package
  \savebox{\myeye}{\eye}%        And direct usage
  \newsavebox{\myeye}
  \newcommand\myanchor{\anchor}% But OK inside command
}

```

3.4 Support for active characters

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

`\initiate@active@char`

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

`\bbl@activate`
`\bbl@deactivate`

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

`\declare@shorthand`

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

`\bbl@add@special`
`\bbl@remove@special`

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

3.5 Support for saving macro definitions

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

`\babel@save`

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

`\babel@savevariable`

A second macro is provided to save the current value of a variable. In this context,

²⁷This mechanism was introduced by Bernd Raichle.

anything that is allowed after the `\the` primitive is considered to be a variable. The macro takes one argument, the *<variable>*.

The effect of the preceding macros is to append a piece of code to the current definition of `\originalTeX`. When `\originalTeX` is expanded, this code restores the previous definition of the control sequence or the previous value of the variable.

3.6 Support for extending macros

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

3.7 Macros common to a number of languages

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

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

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

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

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

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

3.8 Encoding-dependent strings

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

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

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

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

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

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

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

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

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

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

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

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

\EndBabelCommands
```

A real example is:

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

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

²⁸In future releases further categories may be added.

```

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

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

\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\monthviiname{August}
  \SetString\monthixname{September}
  \SetString\monthxname{Oktober}
  \SetString\monthxiname{November}
  \SetString\monthxiiname{Dezenber}
  \SetString\today{\number\day.~%
    \csname month\romannumeral\month name\endcsname\space
    \number\year}

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

\EndBabelCommands

```

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

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

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

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

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

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

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

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

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

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

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

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

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

#1 is replaced by the roman numeral.

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

Sets globally code to be executed at \MakeUppercase and \MakeLowercase. The code would typically be 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 can set for Turkish:

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

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

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

\EndBabelCommands
```

(Note the mapping for OT1 is not complete.)

\SetHyphenMap {<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*).

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

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

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

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

4 Changes

4.1 Changes in babel version 3.9

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

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

Part II

Source code

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

5 Identification and loading of required files

Code documentation is still under revision.

The following description is no longer valid, because switch and plain have been merged into babel.def.

The babel package after unpacking consists of the following files:

switch.def defines macros to set and switch languages.

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

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

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

hyphen.cfg is the file to be used when generating the formats to load hyphenation patterns.

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

6 locale directory

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

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

This is a preliminary documentation.

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

Most keys are self-explanatory.

charset the encoding used in the ini file.

version of the ini file

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

encodings a descriptive list of font encodings.

[captions] section of captions in the file charset

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

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

Keys may be further qualified in a particular language with a suffix starting with an uppercase letter. It can be just a letter (eg, babel.name.A, babel.name.B) or a name (eg, date.long.Nominative, date.long.Formal, but no language is currently using the latter). *Multi-letter* qualifiers are forward compatible in the sense they won’t conflict with new “global” keys (which start always with a lowercase case). There is an exception, however: the section counters has been devised to have arbitrary keys, so you can add lowercased keys if you want.

7 Tools

```
1 <<version=3.52.2228>>
2 <<date=2020/12/21>>
```

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

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

```
3 <<*Basic macros>> ≡
4 \bbl@trace{Basic macros}
5 \def\bbl@stripslash{\expandafter\@gobble\string}
6 \def\bbl@add#1#2{%
7   \bbl@ifunset{\bbl@stripslash#1}%
8   {\def#1{#2}}%
9   {\expandafter\def\expandafter#1\expandafter{#1#2}}
10 \def\bbl@xin@\@expandtwoargs\in@
11 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
12 \def\bbl@cs#1{\csname bbl@#1\endcsname}
13 \def\bbl@c1#1{\csname bbl@#1\language\endcsname}
14 \def\bbl@loop#1#2#3{\bbl@loop#1{#3}#2,\@nnil,}
15 \def\bbl@loopx#1#2{\expandafter\bbl@loop\expandafter#1\expandafter{#2}}
16 \def\bbl@loop#1#2#3,{%
17   \ifx\@nnil#3\relax\else
18     \def#1{#3}#2\bbl@afterfi\bbl@loop#1{#2}%
19   \fi}
20 \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.

```
21 \def\bbl@add@list#1#2{%
22   \edef#1{%
23     \bbl@ifunset{\bbl@stripslash#1}%
24     {}%
25     {\ifx#1\@empty\else#1,\fi}%
26   #2}}
```

`\bbl@afterelse` `\bbl@afterfi` Because the code that is used in the handling of active characters may need to look ahead, we take extra care to ‘throw’ it over the `\else` and `\fi` parts of an `\if`-statement³⁰. These macros will break if another `\if... \fi` statement appears in one of the arguments and it is not enclosed in braces.

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

`\bbl@exp` Now, just syntactical sugar, but it makes partial expansion of some code a lot more simple and readable. Here `\` stands for `\noexpand` and `<.>` for `\noexpand` applied to a built macro name (the latter does not define the macro if undefined to `\relax`, because it is created locally). The result may be followed by extra arguments, if necessary.

```
29 \def\bbl@exp#1{%
30   \begingroup
```

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


```

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

```

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

```

35 \def\bbl@tempa#1{%
36 \long\def\bbl@trim##1##2{%
37 \futurelet\bbl@trim@a\bbl@trim@c##2\@nil\@nil#1\@nil\relax{##1}}%
38 \def\bbl@trim@c{%
39 \ifx\bbl@trim@a\@sptoken
40 \expandafter\bbl@trim@b
41 \else
42 \expandafter\bbl@trim@b\expandafter#1%
43 \fi}%
44 \long\def\bbl@trim@b#1##1 \@nil{\bbl@trim@i##1}}
45 \bbl@tempa{ }
46 \long\def\bbl@trim@i#1\@nil#2\relax#3{#3{#1}}
47 \long\def\bbl@trim@def#1{\bbl@trim{\def#1}}

```

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

```

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

```

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

```

68 \def\bbl@ifblank#1{%
69 \bbl@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
70 \long\def\bbl@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
71 \def\bbl@ifset#1#2#3{%
72 \bbl@ifunset{#1}{#3}{\bbl@exp{\bbl@ifblank{#1}}{#3}{#2}}}

```

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

```

73 \def\bbl@forkv#1#2{%
74   \def\bbl@kvcmd##1##2##3{#2}%
75   \bbl@kvnext#1,\@nil,}
76 \def\bbl@kvnext#1,{%
77   \ifx\@nil#1\relax\else
78     \bbl@ifblank{#1}{\bbl@forkv@eq#1=\@empty=\@nil{#1}}%
79     \expandafter\bbl@kvnext
80   \fi}
81 \def\bbl@forkv@eq#1=#2=#3\@nil#4{%
82   \bbl@trim@def\bbl@forkv@a{#1}%
83   \bbl@trim{\expandafter\bbl@kvcmd\expandafter{\bbl@forkv@a}{#2}{#4}}

```

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

```

84 \def\bbl@vforeach#1#2{%
85   \def\bbl@forcmd##1{#2}%
86   \bbl@fornext#1,\@nil,}
87 \def\bbl@fornext#1,{%
88   \ifx\@nil#1\relax\else
89     \bbl@ifblank{#1}{\bbl@trim\bbl@forcmd{#1}}%
90     \expandafter\bbl@fornext
91   \fi}
92 \def\bbl@foreach#1{\expandafter\bbl@vforeach\expandafter{#1}}

```

\bbl@replace

```

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

```

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

```

105 \ifx\detokenize\undefined\else % Unused macros if old Plain TeX
106   \bbl@exp{\def\\bbl@parsedef##1\detokenize{macro:}}#2->#3\relax{%
107     \def\bbl@tempa{#1}%
108     \def\bbl@tempb{#2}%
109     \def\bbl@tempe{#3}}
110   \def\bbl@sreplace#1#2#3{%
111     \begingroup
112     \expandafter\bbl@parsedef\meaning#1\relax
113     \def\bbl@tempc{#2}%
114     \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
115     \def\bbl@tempd{#3}%

```

```

116 \edef\bbl@tempd{\expandafter\strip@prefix\meaning\bbl@tempd}%
117 \bbl@xin@{\bbl@tempc}{\bbl@tempe}% If not in macro, do nothing
118 \ifin@
119 \bbl@exp{\bbl@replace\bbl@tempe{\bbl@tempc}{\bbl@tempd}}%
120 \def\bbl@tempc{% Expanded an executed below as 'uplevel'
121 \\\makeatletter % "internal" macros with @ are assumed
122 \\\scantokens{%
123 \bbl@tempa\\@namedef{\bbl@stripslash#1}\bbl@tempb{\bbl@tempe}}%
124 \catcode64=\the\catcode64\relax}% Restore @
125 \else
126 \let\bbl@tempc\@empty % Not \relax
127 \fi
128 \bbl@exp{% For the 'uplevel' assignments
129 \endgroup
130 \bbl@tempc}} % empty or expand to set #1 with changes
131 \fi

```

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

```

132 \def\bbl@ifsamestring#1#2{%
133 \begingroup
134 \protected@edef\bbl@tempb{#1}%
135 \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
136 \protected@edef\bbl@tempc{#2}%
137 \edef\bbl@tempc{\expandafter\strip@prefix\meaning\bbl@tempc}%
138 \ifx\bbl@tempb\bbl@tempc
139 \aftergroup\@firstoftwo
140 \else
141 \aftergroup\@secondoftwo
142 \fi
143 \endgroup}
144 \chardef\bbl@engine=%
145 \ifx\directlua\@undefined
146 \ifx\XeTeXinputencoding\@undefined
147 \z@
148 \else
149 \tw@
150 \fi
151 \else
152 \@ne
153 \fi

```

A somewhat hackish tool (hence its name) to avoid spurious spaces in some contexts.

```

154 \def\bbl@bsphack{%
155 \ifhmode
156 \hskip\z@skip
157 \def\bbl@esphack{\loop\ifdim\lastskip>\z@\unskip\repeat\unskip}%
158 \else
159 \let\bbl@esphack\@empty
160 \fi}

```

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

```

161 \def\bbl@cased{%
162 \ifx\oe\OE

```

```

163 \expandafter\in@\expandafter
164 {\expandafter\OE\expandafter}\expandafter{\oe}%
165 \ifin@
166 \bbl@afterelse\expandafter\MakeUppercase
167 \else
168 \bbl@afterfi\expandafter\MakeLowercase
169 \fi
170 \else
171 \expandafter\@firstofone
172 \fi}
173 <</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 .

```

174 <<*Make sure ProvidesFile is defined>> ≡
175 \ifx\ProvidesFile\@undefined
176 \def\ProvidesFile#1[#2 #3 #4]{%
177 \wlog{File: #1 #4 #3 <#2>}%
178 \let\ProvidesFile\@undefined}
179 \fi
180 <</Make sure ProvidesFile is defined>>

```

7.1 Multiple languages

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

```

181 <<*Define core switching macros>> ≡
182 \ifx\language\@undefined
183 \csname newcount\endcsname\language
184 \fi
185 <</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` This macro was introduced for \TeX < 2. Preserved for compatibility.

```

186 <<*Define core switching macros>> ≡
187 <<*Define core switching macros>> ≡
188 \countdef\last@language=19 % TODO. why? remove?
189 \def\addlanguage{\csname newlanguage\endcsname}
190 <</Define core switching macros>>

```

Now we make sure all required files are loaded. When the command `\AtBeginDocument` doesn't exist we assume that we are dealing with a plain-based format or \LaTeX 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.

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

This file also takes care of a number of compatibility issues with other packages and defines a few additional package options. Apart from all the language options below we also have a few options that influence the behavior of language definition files.

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

The first two options are for debugging.

```
191 <{*package}>
192 \NeedsTeXFormat{LaTeX2e}[2005/12/01]
193 \ProvidesPackage{babel}[<<date>> <<version>> The Babel package]
194 \@ifpackagewith{babel}{debug}
195   {\providecommand\bbl@trace[1]{\message{^^J[ #1 ]}}%
196    \let\bbl@debug\@firstofone
197    \ifx\directlua\undefined\else
198      \directlua{ Babel = Babel or {}
199        Babel.debug = true }%
200    \fi}
201 {\providecommand\bbl@trace[1]{}%
202  \let\bbl@debug\gobble
203  \ifx\directlua\undefined\else
204    \directlua{ Babel = Babel or {}
205      Babel.debug = false }%
206  \fi}
207 <<Basic macros>>
208 % Temporarily repeat here the code for errors
209 \def\bbl@error#1#2{%
210   \begingroup
211     \def\{\MessageBreak}%
212     \PackageError{babel}{#1}{#2}%
213   \endgroup}
214 \def\bbl@warning#1{%
215   \begingroup
216     \def\{\MessageBreak}%
217     \PackageWarning{babel}{#1}%
218   \endgroup}
219 \def\bbl@infowarn#1{%
220   \begingroup
221     \def\{\MessageBreak}%
222     \GenericWarning
223       {(babel) \@spaces\@spaces\@spaces}%
224       {Package babel Info: #1}%
225   \endgroup}
226 \def\bbl@info#1{%
227   \begingroup
228     \def\{\MessageBreak}%
229     \PackageInfo{babel}{#1}%
230   \endgroup}
231 \def\bbl@nocaption{\protect\bbl@nocaption@i}
232 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
233   \global\@namedef{#2}{\textbf{?#1?}}%
234   \@nameuse{#2}%
235   \bbl@warning{%
236     \@backslashchar#2 not set. Please, define it\\%
237     after the language has been loaded (typically\\%
238     in the preamble) with something like:\\%
239     \string\renewcommand\@backslashchar#2{..}\\%
```

```

240   Reported}}
241 \def\bbl@tentative{\protect\bbl@tentative@i}
242 \def\bbl@tentative@i#1{%
243   \bbl@warning{%
244     Some functions for '#1' are tentative.\\%
245     They might not work as expected and their behavior\\%
246     may change in the future.\\%
247     Reported}}
248 \def\@nolanerr#1{%
249   \bbl@error
250   {You haven't defined the language #1\space yet.\\%
251     Perhaps you misspelled it or your installation\\%
252     is not complete}%
253   {Your command will be ignored, type <return> to proceed}}
254 \def\@nopatterns#1{%
255   \bbl@warning
256   {No hyphenation patterns were preloaded for\\%
257     the language `#1' into the format.\\%
258     Please, configure your TeX system to add them and\\%
259     rebuild the format. Now I will use the patterns\\%
260     preloaded for \bbl@nulllanguage\space instead}}
261   % End of errors
262 \@ifpackagewith{babel}{silent}
263 {\let\bbl@info\@gobble
264   \let\bbl@infowarn\@gobble
265   \let\bbl@warning\@gobble}
266 {}
267 %
268 \def\AfterBabelLanguage#1{%
269   \global\expandafter\bbl@add\csname#1.ldf-h@@k\endcsname}%

```

If the format created a list of loaded languages (in \bbl@languages), get the name of the 0-th to show the actual language used. Also available with base, because it just shows info.

```

270 \ifx\bbl@languages\undefined\else
271   \begingroup
272     \catcode`\^^I=12
273     \@ifpackagewith{babel}{showlanguages}{%
274       \begingroup
275         \def\bbl@elt#1#2#3#4{\wlog{#2^^I#1^^I#3^^I#4}}%
276         \wlog{<*languages>}%
277         \bbl@languages
278         \wlog{</languages>}%
279       \endgroup}{%
280     \endgroup
281     \def\bbl@elt#1#2#3#4{%
282       \ifnum#2=\z@
283         \gdef\bbl@nulllanguage{#1}%
284         \def\bbl@elt##1##2##3##4{}%
285       \fi}%
286     \bbl@languages
287 \fi%

```

7.3 base

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

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

```

288 \bbl@trace{Defining option 'base'}
289 \@ifpackagewith{babel}{base}{%
290   \let\bbl@onlyswitch\@empty
291   \let\bbl@provide@locale\relax
292   \input babel.def
293   \let\bbl@onlyswitch\@undefined
294   \ifx\directlua\@undefined
295     \DeclareOption*{\bbl@patterns{\CurrentOption}}%
296   \else
297     \input luababel.def
298     \DeclareOption*{\bbl@patterns@lua{\CurrentOption}}%
299   \fi
300   \DeclareOption{base}{}%
301   \DeclareOption{showlanguages}{}%
302   \ProcessOptions
303   \global\expandafter\let\csname opt@babel.sty\endcsname\relax
304   \global\expandafter\let\csname ver@babel.sty\endcsname\relax
305   \global\let\@ifl@ter@@\@ifl@ter
306   \def\@ifl@ter#1#2#3#4#5{\global\let\@ifl@ter\@ifl@ter@@}%
307   \endinput}{}%
308 % \end{macrocode}
309 %
310 % \subsection{\texttt{key=value} options and other general option}
311 %
312 %   The following macros extract language modifiers, and only real
313 %   package options are kept in the option list. Modifiers are saved
314 %   and assigned to |\BabelModifiers| at |\bbl@load@language|; when
315 %   no modifiers have been given, the former is |\relax|. How
316 %   modifiers are handled are left to language styles; they can use
317 %   |\in@|, loop them with |\@for| or load |keyval|, for example.
318 %
319 %   \begin{macrocode}
320 \bbl@trace{key=value and another general options}
321 \bbl@csarg\let\tempa\expandafter\csname opt@babel.sty\endcsname
322 \def\bbl@tempb#1.#2{% Remove trailing dot
323   #1\ifx\@empty#2\else,\bbl@afterfi\bbl@tempb#2\fi}%
324 \def\bbl@tempd#1.#2\@nnil{% TODO. Refactor lists?
325   \ifx\@empty#2%
326     \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
327   \else
328     \in@{,provide,},{, #1,}%
329     \ifin@
330       \edef\bbl@tempc{%
331         \ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.\bbl@tempb#2}%
332     \else
333       \in@{=}{, #1}%
334       \ifin@
335         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1.#2}%
336       \else
337         \edef\bbl@tempc{\ifx\bbl@tempc\@empty\else\bbl@tempc,\fi#1}%
338       \bbl@csarg\edef{mod#1}{\bbl@tempb#2}%
339     \fi
340   \fi
341 \fi}
342 \let\bbl@tempc\@empty
343 \bbl@foreach\bbl@tempa{\bbl@tempd#1.\@empty\@nnil}

```

```
344 \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.

```
345 \DeclareOption{KeepShorthandsActive}{}
346 \DeclareOption{activeacute}{}
347 \DeclareOption{activegrave}{}
348 \DeclareOption{debug}{}
349 \DeclareOption{noconfigs}{}
350 \DeclareOption{showlanguages}{}
351 \DeclareOption{silent}{}
352 \DeclareOption{mono}{}
353 \DeclareOption{shorthands=off}{\bbl@tempa shorthands=\bbl@tempa}
354 \chardef\bbl@iniflag\z@
355 \DeclareOption{provide=*}{\chardef\bbl@iniflag\@ne} % main -> +1
356 \DeclareOption{provide+=*}{\chardef\bbl@iniflag\@tw@} % add = 2
357 \DeclareOption{provide*=*}{\chardef\bbl@iniflag\@thr@} % add + main
358 % A separate option
359 \let\bbl@autoload@options\@empty
360 \DeclareOption{provide@=*}{\def\bbl@autoload@options{import}}
361 % Don't use. Experimental. TODO.
362 \newif\ifbbl@single
363 \DeclareOption{selectors=off}{\bbl@singletrue}
364 <<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.

```
365 \let\bbl@opt@shorthands\@nnil
366 \let\bbl@opt@config\@nnil
367 \let\bbl@opt@main\@nnil
368 \let\bbl@opt@headfoot\@nnil
369 \let\bbl@opt@layout\@nnil
```

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

```
370 \def\bbl@tempa#1=#2\bbl@tempa{%
371   \bbl@csarg\ifx{opt@#1}\@nnil
372     \bbl@csarg\edef{opt@#1}{#2}%
373   \else
374     \bbl@error
375     {Bad option `#1=#2'. Either you have misspelled the\\%
376     key or there is a previous setting of `#1'. Valid\\%
377     keys are, among others, `shorthands', `main', `bidi',\\%
378     `strings', `config', `headfoot', `safe', `math'.}%
379     {See the manual for further details.}
380   \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.

```
381 \let\bbl@language@opts\@empty
382 \DeclareOption*{%
383   \bbl@xin@{\string=}{\CurrentOption}%
```



```

384 \ifin@
385 \expandafter\bbbl@tempa\CurrentOption\bbbl@tempa
386 \else
387 \bbbl@add@list\bbbl@language@opts{\CurrentOption}%
388 \fi}

```

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

```

389 \ProcessOptions*

```

7.4 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 `\bbbl@ifshorthand` is always true, and it is always false if `shorthands` is empty. Also, some code makes sense only with `shorthands=...`

```

390 \bbbl@trace{Conditional loading of shorthands}
391 \def\bbbl@sh@string#1{%
392   \ifx#1\@empty\else
393     \ifx#1t\string~%
394     \else\ifx#1c\string,%
395     \else\string#1%
396     \fi\fi
397   \expandafter\bbbl@sh@string
398   \fi}
399 \ifx\bbbl@opt@shorthands\@nnil
400   \def\bbbl@ifshorthand#1#2#3{#2}%
401 \else\ifx\bbbl@opt@shorthands\@empty
402   \def\bbbl@ifshorthand#1#2#3{#3}%
403 \else

```

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

```

404 \def\bbbl@ifshorthand#1{%
405   \bbbl@xin@\string#1}{\bbbl@opt@shorthands}%
406   \ifin@
407     \expandafter\@firstoftwo
408   \else
409     \expandafter\@secondoftwo
410   \fi}

```

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

```

411 \edef\bbbl@opt@shorthands{%
412   \expandafter\bbbl@sh@string\bbbl@opt@shorthands\@empty}%

```

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

```

413 \bbbl@ifshorthand{'}%
414   {\PassOptionsToPackage{activeacute}{babel}}{}
415 \bbbl@ifshorthand{`}%
416   {\PassOptionsToPackage{activegrave}{babel}}{}
417 \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.

```

418 \ifx\bbbl@opt@headfoot\@nnil\else

```

```

419 \g@addto@macro\@resetactivechars{%
420   \set@typeset@protect
421   \expandafter\select@language@x\expandafter{\bbl@opt@headfoot}%
422   \let\protect\noexpand}
423 \fi

```

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

```

424 \ifx\bbl@opt@safe\undefined
425   \def\bbl@opt@safe{BR}
426 \fi
427 \ifx\bbl@opt@main\@nnil\else
428   \edef\bbl@language@opts{%
429     \ifx\bbl@language@opts\empty\else\bbl@language@opts,\fi
430     \bbl@opt@main}
431 \fi

```

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

```

432 \bbl@trace{Defining IfBabelLayout}
433 \ifx\bbl@opt@layout\@nnil
434   \newcommand\IfBabelLayout[3]{#3}%
435 \else
436   \newcommand\IfBabelLayout[1]{%
437     \@expandtwoargs\in@{.#1.}{.\bbl@opt@layout.}%
438     \ifin@
439       \expandafter\@firstoftwo
440     \else
441       \expandafter\@secondoftwo
442     \fi}
443 \fi

```

Common definitions. *In progress.* Still based on `babel.def`, but the code should be moved here.

```

444 \input babel.def

```

7.5 Cross referencing macros

The \LaTeX book states:

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

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

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

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

```

445 <<{*More package options}>> ≡
446 \DeclareOption{safe=none}{\let\bbl@opt@safe\empty}
447 \DeclareOption{safe=bib}{\def\bbl@opt@safe{B}}
448 \DeclareOption{safe=ref}{\def\bbl@opt@safe{R}}
449 <</More package options>>

```

`\@newl@bel` 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.

```

450 \bbl@trace{Cross referencing macros}
451 \ifx\bbl@opt@safe\empty\else
452   \def\@newl@bel#1#2#3{%
453     {\@safe@activetrue
454       \bbl@ifunset{#1@#2}%
455       \relax
456       {\gdef\@multiplelabels{%
457         \@latex@warning@no@line{There were multiply-defined labels}}}%
458       \@latex@warning@no@line{Label `#2' multiply defined}}}%
459   \global\@namedef{#1@#2}{#3}}

```

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

```

460 \CheckCommand*\@testdef[3]{%
461   \def\reserved@a{#3}%
462   \expandafter\ifx\csname#1@#2\endcsname\reserved@a
463   \else
464     \@tempswatrue
465   \fi}

```

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

```

466 \def\@testdef#1#2#3{% TODO. With @samestring?
467   \@safe@activetrue
468   \expandafter\let\expandafter\bbl@tempa\csname #1@#2\endcsname
469   \def\bbl@tempb{#3}%
470   \@safe@activetrue
471   \ifx\bbl@tempa\relax
472   \else
473     \edef\bbl@tempa{\expandafter\strip@prefix\meaning\bbl@tempa}%
474   \fi
475   \edef\bbl@tempb{\expandafter\strip@prefix\meaning\bbl@tempb}%
476   \ifx\bbl@tempa\bbl@tempb
477   \else
478     \@tempswatrue
479   \fi}
480 \fi

```

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

```

481 \bbl@xin@{R}\bbl@opt@safe
482 \ifin@
483   \bbl@redefineroobust\ref#1{%
484     \@safe@activetrue\org@ref{#1}\@safe@activetrue}
485   \bbl@redefineroobust\pageref#1{%
486     \@safe@activetrue\org@pageref{#1}\@safe@activetrue}
487 \else
488   \let\org@ref\ref
489   \let\org@pageref\pageref
490 \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.

```
491 \bbl@xin@{B}\bbl@opt@safe
492 \ifin@
493 \bbl@redefine\@citex[#1]#2{%
494   \@safe@activetrue\edef\@tempa{#2}\@safe@activesfalse
495   \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.

```
496 \AtBeginDocument{%
497   \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.)

```
498   \def\@citex[#1][#2]#3{%
499     \@safe@activetrue\edef\@tempa{#3}\@safe@activesfalse
500     \org@@citex[#1][#2]{\@tempa}}%
501   }{}}
```

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

```
502 \AtBeginDocument{%
503   \ifpackageloaded{cite}{%
504     \def\@citex[#1]#2{%
505       \@safe@activetrue\org@@citex[#1]{#2}\@safe@activesfalse}%
506     }{}}
```

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

```
507 \bbl@redefine\nocite#1{%
508   \@safe@activetrue\org@nocite{#1}\@safe@activesfalse}
```

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

```
509 \bbl@redefine\bibcite{%
510   \bbl@cite@choice
511   \bibcite}
```

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

```
512 \def\bbl@bibcite#1#2{%
513   \org@bibcite{#1}{\@safe@activesfalse#2}}
```

`\bbl@cite@choice` The macro `\bbl@cite@choice` determines which definition of `\bibcite` is needed. First we give `\bibcite` its default definition.

```
514 \def\bbl@cite@choice{%
515   \global\let\bibcite\bbl@bibcite
516   \@ifpackageloaded{natbib}{\global\let\bibcite\org@bibcite}{}%
517   \@ifpackageloaded{cite}{\global\let\bibcite\org@bibcite}{}%
518   \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.

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

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

```
520 \bbl@redefine\@bibitem#1{%
521   \@safe@activetrue\org@bibitem{#1}\@safe@activesfalse}
522 \else
523   \let\org@nocite\nocite
524   \let\org@@citex\@citex
525   \let\org@bibcite\bibcite
526   \let\org@@bibitem\@bibitem
527 \fi
```

7.6 Marks

`\markright` Because the output routine is asynchronous, we must pass the current language attribute to the head lines. To achieve this we need to adapt the definition of `\markright` and `\markboth` somewhat. However, headlines and footlines can contain text outside marks; for that we must take some actions in the output routine if the 'headfoot' options is used. We need to make some redefinitions to the output routine to avoid an endless loop and to correctly handle the page number in bidi documents.

```
528 \bbl@trace{Marks}
529 \IfBabelLayout{sectioning}
530   {\ifx\bbl@opt@headfoot\@nnil
531     \g@addto@macro\@resetactivechars{%
532       \set@typeset@protect
533       \expandafter\select@language@x\expandafter{\bbl@main@language}%
534       \let\protect\noexpand
535       \ifcase\bbl@bidimode\else % Only with bidi. See also above
536         \edef\thepage{%
537           \noexpand\babelsublr{\unexpanded\expandafter{\thepage}}}%
538       \fi}%
539   \fi}
540 {\ifbbl@single\else
541   \bbl@ifunset{markright }{\bbl@redefine\bbl@redefineroobust
542     \markright#1{%
543       \bbl@ifblank{#1}%
544       {\org@markright}{}}%
545     {\toks@{#1}%
546       \bbl@exp{%
547         \org@markright{\protect\foreignlanguage{\language}\thepage}%
548         {\protect\bbl@restore@actives\the\toks@}}}%
549   }
```

`\markboth` The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The documentclasses `report` and `book` define and set the headings for the page. While doing so they also store a copy of `\markboth` in `\@mkboth`. Therefore we need

to check whether `\mkboth` has already been set. If so we need to do that again with the new definition of `\markboth`. (As of Oct 2019, \TeX stores the definition in an intermediate macro, so it's not necessary anymore, but it's preserved for older versions.)

```

549 \ifx\mkboth\markboth
550   \def\bbl@tempc{\let\mkboth\markboth}
551 \else
552   \def\bbl@tempc{}
553 \fi
554 \bbl@ifunset{markboth }{\bbl@redefine\bbl@redefineroobust
555 \markboth#1#2}%
556   \protected@edef\bbl@tempb##1{%
557     \protect\foreignlanguage
558     {\language\name}{\protect\bbl@restore@actives##1}}%
559   \bbl@ifblank{#1}%
560     {\toks@{}}%
561     {\toks@\expandafter{\bbl@tempb{#1}}}%
562   \bbl@ifblank{#2}%
563     {\@temptokena{}}%
564     {\@temptokena\expandafter{\bbl@tempb{#2}}}%
565   \bbl@exp{\org@markboth{\the\toks@}{\the\@temptokena}}
566   \bbl@tempc
567 \fi} % end ifbbl@single, end \IfBabelLayout

```

7.7 Preventing clashes with other packages

7.7.1 `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.

We want to revert the definition of `\pageref` and `\ref` to their original definition for the first argument of `\ifthenelse`, so we first need to store their current meanings. Then we can set the `\@safe@actives` switch and call the original `\ifthenelse`. In order to be able to use shorthands in the second and third arguments of `\ifthenelse` the resetting of the switch *and* the definition of `\pageref` happens inside those arguments.

```

568 \bbl@trace{Preventing clashes with other packages}
569 \bbl@xin@{R}\bbl@opt@safe
570 \ifin@
571 \AtBeginDocument{%
572   \@ifpackageloaded{ifthen}{%
573     \bbl@redefine@long\ifthenelse#1#2#3{%
574       \let\bbl@temp@pref\pageref
575       \let\pageref\org@pageref
576       \let\bbl@temp@ref\ref
577       \let\ref\org@ref
578       \@safe@activestrue
579       \org@ifthenelse{#1}%
580       {\let\pageref\bbl@temp@pref
581        \let\ref\bbl@temp@ref

```

```

582         \@safe@activesfalse
583         #2}%
584     {\let\pageref\bb1@temp@pref
585      \let\ref\bb1@temp@ref
586      \@safe@activesfalse
587      #3}%
588     }%
589   }{}%
590 }

```

7.7.2 varioref

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

```

591 \AtBeginDocument{%
592   \ifpackageloaded{varioref}{%
593     \bb1@redefine\@@vpageref#1[#2]#3{%
594       \@safe@activetrue
595       \org@@vpageref{#1}[#2]{#3}%
596       \@safe@activesfalse}%
597     \bb1@redefine\vrefpagenum#1#2{%
598       \@safe@activetrue
599       \org\vrefpagenum{#1}{#2}%
600       \@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.

```

601   \expandafter\def\csname Ref \endcsname#1{%
602     \protected@edef\@tempa{\org@ref{#1}}\expandafter\MakeUppercase\@tempa}
603   }{}%
604 }
605 \fi

```

7.7.3 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. 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.

```

606 \AtEndOfPackage{%
607   \AtBeginDocument{%
608     \ifpackageloaded{hhline}%
609     {\expandafter\ifx\csname normal@char\string\endcsname\relax
610      \else
611        \makeatletter
612        \def\@currname{hhline}\input{hhline.sty}\makeatother
613        \fi}%
614     {}}}

```

7.7.4 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. TODO. Still true? Commented out in 2020/07/27.

```
615 % \AtBeginDocument{%
616 %   \ifx\pdfstringdefDisableCommands\@undefined\else
617 %     \pdfstringdefDisableCommands{\languageshorthands{system}}%
618 %   \fi}
```

7.7.5 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.

```
619 \DeclareRobustCommand{\FOREIGNLANGUAGE}[1]{%
620   \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. This command is deprecated. Use the tools provided by L^AT_EX.

```
621 \def\substitutefontfamily#1#2#3{%
622   \lowercase{\immediate\openout15=#1#2.fd\relax}%
623   \immediate\write15{%
624     \string\ProvidesFile{#1#2.fd}%
625     [\the\year/\two@digits{\the\month}/\two@digits{\the\day}
626     \space generated font description file]^{}
627     \string\DeclareFontFamily{#1}{#2}{}^{}
628     \string\DeclareFontShape{#1}{#2}{m}{n}{<->ssub * #3/m/n}{}^{}
629     \string\DeclareFontShape{#1}{#2}{m}{it}{<->ssub * #3/m/it}{}^{}
630     \string\DeclareFontShape{#1}{#2}{m}{sl}{<->ssub * #3/m/sl}{}^{}
631     \string\DeclareFontShape{#1}{#2}{m}{sc}{<->ssub * #3/m/sc}{}^{}
632     \string\DeclareFontShape{#1}{#2}{b}{n}{<->ssub * #3/bx/n}{}^{}
633     \string\DeclareFontShape{#1}{#2}{b}{it}{<->ssub * #3/bx/it}{}^{}
634     \string\DeclareFontShape{#1}{#2}{b}{sl}{<->ssub * #3/bx/sl}{}^{}
635     \string\DeclareFontShape{#1}{#2}{b}{sc}{<->ssub * #3/bx/sc}{}^{}
636   }%
637   \closeout15
638 }
639 \@onlypreamble\substitutefontfamily
```

7.8 Encoding and fonts

Because documents may use non-ASCII font encodings, we make sure that the logos of T_EX and L^AT_EX 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`

```
640 \bbl@trace{Encoding and fonts}
```



```

641 \newcommand\BabelNonASCII{LGR,X2,OT2,OT3,OT6,LHE,LWN,LMA,LMC,LMS,LMU,PU,PD1}
642 \newcommand\BabelNonText{TS1,T3,TS3}
643 \let\org@TeX\TeX
644 \let\org@LaTeX\LaTeX
645 \let\ensureascii\@firstofone
646 \AtBeginDocument{%
647   \in@false
648   \bbl@foreach\BabelNonASCII{% is there a text non-ascii enc?
649     \ifin@false
650       \lowercase{\bbl@xin@{,#1enc.def,},{,\@filelist,}}%
651     \fi}%
652   \ifin@ % if a text non-ascii has been loaded
653     \def\ensureascii#1{{\fontencoding{OT1}\selectfont#1}}%
654     \DeclareTextCommandDefault{\TeX}{\org@TeX}%
655     \DeclareTextCommandDefault{\LaTeX}{\org@LaTeX}%
656     \def\bbl@tempb#1\@{\uppercase{\bbl@tempc#1}ENC.DEF\@empty\@}%
657     \def\bbl@tempc#1ENC.DEF#2\@{\%
658       \ifx\@empty#2\else
659         \bbl@ifunset{T@#1}%
660         {}%
661         {\bbl@xin@{,#1,},{,\BabelNonASCII,\BabelNonText,}}%
662         \ifin@
663           \DeclareTextCommand{\TeX}{#1}{\ensureascii{\org@TeX}}%
664           \DeclareTextCommand{\LaTeX}{#1}{\ensureascii{\org@LaTeX}}%
665         \else
666           \def\ensureascii##1{{\fontencoding{#1}\selectfont##1}}%
667         \fi}%
668     \fi}%
669   \bbl@foreach\@filelist{\bbl@tempb#1\@}% TODO - \@ de mas??
670   \bbl@xin@{\cf@encoding,}{,\BabelNonASCII,\BabelNonText,}%
671   \ifin@false
672     \edef\ensureascii#1{%
673       \noexpand\fontencoding{\cf@encoding}\noexpand\selectfont#1}}%
674   \fi
675 \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.

```

676 \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.

```

677 \AtBeginDocument{%
678   \@ifpackageloaded{fontspec}%
679   {\xdef\latinencoding{%
680     \ifx\UTFencname\undefined
681       EU\ifcase\bbl@engine\or2\or1\fi
682     \else
683       \UTFencname
684     \fi}}%
685   {\gdef\latinencoding{OT1}}%

```

```

686 \ifx\cf@encoding\bbl@t@one
687 \xdef\latinencoding{\bbl@t@one}%
688 \else
689 \ifx\@fontenc@load@list\@undefined
690 \@ifl@aded{def}{t1enc}{\xdef\latinencoding{\bbl@t@one}}{}%
691 \else
692 \def\@elt#1{, #1,}%
693 \edef\bbl@tempa{\expandafter\@gobbletwo\@fontenc@load@list}%
694 \let\@elt\relax
695 \bbl@xin@{, T1, }\bbl@tempa
696 \ifin@
697 \xdef\latinencoding{\bbl@t@one}%
698 \fi
699 \fi
700 \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.

```

701 \DeclareRobustCommand{\latintext}{%
702 \fontencoding{\latinencoding}\selectfont
703 \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.

```

704 \ifx\@undefined\DeclareTextFontCommand
705 \DeclareRobustCommand{\textlatin}[1]{\leavevmode{\latintext #1}}
706 \else
707 \DeclareTextFontCommand{\textlatin}{\latintext}
708 \fi

```

7.9 Basic bidi support

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

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

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

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

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

```

709 \ifodd\bbl@engine
710   \def\bbl@activate@preotf{%
711     \let\bbl@activate@preotf\relax % only once
712     \directlua{
713       Babel = Babel or {}
714       %
715       function Babel.pre_otfload_v(head)
716         if Babel.numbers and Babel.digits_mapped then
717           head = Babel.numbers(head)
718         end
719         if Babel.bidi_enabled then
720           head = Babel.bidi(head, false, dir)
721         end
722         return head
723       end
724       %
725       function Babel.pre_otfload_h(head, gc, sz, pt, dir)
726         if Babel.numbers and Babel.digits_mapped then
727           head = Babel.numbers(head)
728         end
729         if Babel.bidi_enabled then
730           head = Babel.bidi(head, false, dir)
731         end
732         return head
733       end
734       %
735       luatexbase.add_to_callback('pre_linebreak_filter',
736         Babel.pre_otfload_v,
737         'Babel.pre_otfload_v',
738       luatexbase.priority_in_callback('pre_linebreak_filter',
739         'luaotfload.node_processor') or nil)
740       %
741       luatexbase.add_to_callback('hpack_filter',
742         Babel.pre_otfload_h,
743         'Babel.pre_otfload_h',
744       luatexbase.priority_in_callback('hpack_filter',
745         'luaotfload.node_processor') or nil)
746     }}
747 \fi

```

The basic setup. In `luatex`, the output is modified at a very low level to set the `\bodydir` to the `\pagedir`.

```

748 \bbl@trace{Loading basic (internal) bidi support}
749 \ifodd\bbl@engine
750   \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
751     \let\bbl@beforeforeign\leavevmode
752     \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
753     \RequirePackage{luatexbase}
754     \bbl@activate@preotf
755     \directlua{
756       require('babel-data-bidi.lua')
757       \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
758         require('babel-bidi-basic.lua')
759       \or
760         require('babel-bidi-basic-r.lua')

```

```

761     \fi}
762 % TODO - to locale_props, not as separate attribute
763 \newattribute\bbl@attr@dir
764 % TODO. I don't like it, hackish:
765 \bbl@exp{\output{\bodydir\pagedir\the\output}}
766 \AtEndOfPackage{\EnableBabelHook{babel-bidi}}
767 \fi\fi
768 \else
769 \ifnum\bbl@bidimode>100 \ifnum\bbl@bidimode<200
770   \bbl@error
771   {The bidi method 'basic' is available only in\\%
772     luatex. I'll continue with 'bidi=default', so\\%
773     expect wrong results}%
774   {See the manual for further details.}%
775   \let\bbl@beforeforeign\leavevmode
776   \AtEndOfPackage{%
777     \EnableBabelHook{babel-bidi}%
778     \bbl@xebidipar}
779 \fi\fi
780 \def\bbl@loadxebidi#1{%
781   \ifx\RTLfootnotetext\@undefined
782     \AtEndOfPackage{%
783       \EnableBabelHook{babel-bidi}%
784       \ifx\fontspec\@undefined
785         \bbl@loadfontspec % bidi needs fontspec
786         \fi
787       \usepackage#1{bidi}}%
788   \fi}
789 \ifnum\bbl@bidimode>200
790   \ifcase\expandafter\@gobbletwo\the\bbl@bidimode\or
791     \bbl@tentative{bidi=bidi}
792     \bbl@loadxebidi{}
793   \or
794     \bbl@loadxebidi{[rldocument]}
795   \or
796     \bbl@loadxebidi{}
797   \fi
798 \fi
799 \fi
800 \ifnum\bbl@bidimode=\@ne
801   \let\bbl@beforeforeign\leavevmode
802   \ifodd\bbl@engine
803     \newattribute\bbl@attr@dir
804     \bbl@exp{\output{\bodydir\pagedir\the\output}}%
805   \fi
806   \AtEndOfPackage{%
807     \EnableBabelHook{babel-bidi}%
808     \ifodd\bbl@engine\else
809       \bbl@xebidipar
810     \fi}
811 \fi

```

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

```

812 \bbl@trace{Macros to switch the text direction}
813 \def\bbl@alscripts{,Arabic,Syriac,Thaana,}
814 \def\bbl@rscripts{% TODO. Base on codes ??
815   ,Imperial Aramaic,Avestan,Cypriot,Hatran,Hebrew,%
816   Old Hungarian,Old Hungarian,Lydian,Mandaean,Manichaeen,%

```

```

817 Manichaeae, Meroitic Cursive, Meroitic, Old North Arabian, %
818 Nabataean, N'Ko, Orkhon, Palmyrene, Inscriptional Pahlavi, %
819 Psalter Pahlavi, Phoenician, Inscriptional Parthian, Samaritan, %
820 Old South Arabian, %}
821 \def\babelprovide@dirs#1{%
822   \babel@x in@{\csname babel@name@#1\endcsname}{\babel@alscripts\babel@rscripts}%
823   \ifin@
824     \global\babel@csarg\chardef{wdir@#1}\@ne
825     \babel@x in@{\csname babel@name@#1\endcsname}{\babel@alscripts}%
826     \ifin@
827       \global\babel@csarg\chardef{wdir@#1}\tw@ % useless in xetex
828     \fi
829   \else
830     \global\babel@csarg\chardef{wdir@#1}\z@
831   \fi
832   \ifodd\babel@engine
833     \babel@csarg\ifcase{wdir@#1}%
834       \directlua{ Babel.locale_props[\the\localeid].texdir = 'l' }%
835     \or
836       \directlua{ Babel.locale_props[\the\localeid].texdir = 'r' }%
837     \or
838       \directlua{ Babel.locale_props[\the\localeid].texdir = 'al' }%
839     \fi
840   \fi}
841 \def\babel@switchdir{%
842   \babel@ifunset{babel@sys@\languagename}{\babel@provide@sys{\languagename}}{}%
843   \babel@ifunset{babel@wdir@\languagename}{\babel@provide@dirs{\languagename}}{}%
844   \babel@exp{\babel@setdirs\babel@cl{wdir}}%
845 \def\babel@setdirs#1{% TODO - math
846   \ifcase\babel@select@type % TODO - strictly, not the right test
847     \babel@bodydir{#1}%
848     \babel@pardir{#1}%
849   \fi
850   \babel@texdir{#1}}
851 % TODO. Only if \babel@bidimode > 0?:
852 \AddBabelHook{babel-bidi}{afterextras}{\babel@switchdir}
853 \DisableBabelHook{babel-bidi}

```

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

```

854 \ifodd\babel@engine % luatex=1
855   \chardef\babel@thetexdir\z@
856   \chardef\babel@thepardir\z@
857   \def\babel@getluadir#1{%
858     \directlua{
859       if tex.#1dir == 'TLT' then
860         tex.sprint('0')
861       elseif tex.#1dir == 'TRT' then
862         tex.sprint('1')
863       end}}
864   \def\babel@setluadir#1#2#3{% 1=text/par.. 2=\texdir.. 3=0 lr/1 r1
865     \ifcase#3\relax
866       \ifcase\babel@getluadir{#1}\relax\else
867         #2 TLT\relax
868       \fi
869     \else
870       \ifcase\babel@getluadir{#1}\relax
871         #2 TRT\relax
872       \fi
873     \fi}

```

```

874 \def\bbl@textdir#1{%
875   \bbl@setluadir{text}\textdir{#1}%
876   \chardef\bbl@thetextdir#1\relax
877   \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
878 \def\bbl@pardir#1{%
879   \bbl@setluadir{par}\pardir{#1}%
880   \chardef\bbl@thepardir#1\relax}
881 \def\bbl@bodydir{\bbl@setluadir{body}\bodydir}
882 \def\bbl@pagedir{\bbl@setluadir{page}\pagedir}
883 \def\bbl@dirparastext{\pardir\the\textdir\relax}%   %%%
884 % Sadly, we have to deal with boxes in math with basic.
885 % Activated every math with the package option bidi=:
886 \def\bbl@mathboxdir{%
887   \ifcase\bbl@thetextdir\relax
888     \everyhbox{\textdir TLT\relax}%
889   \else
890     \everyhbox{\textdir TRT\relax}%
891   \fi}
892 \frozen@everymath\expandafter{%
893   \expandafter\bbl@mathboxdir\the\frozen@everymath}
894 \frozen@everydisplay\expandafter{%
895   \expandafter\bbl@mathboxdir\the\frozen@everydisplay}
896 \else % pdftex=0, xetex=2
897   \newcount\bbl@dirlevel
898   \chardef\bbl@thetextdir\z@
899   \chardef\bbl@thepardir\z@
900   \def\bbl@textdir#1{%
901     \ifcase#1\relax
902       \chardef\bbl@thetextdir\z@
903       \bbl@textdir@i\beginL\endL
904     \else
905       \chardef\bbl@thetextdir\@ne
906       \bbl@textdir@i\beginR\endR
907     \fi}
908   \def\bbl@textdir@i#1#2{%
909     \ifhmode
910       \ifnum\currentgrouplevel>\z@
911         \ifnum\currentgrouplevel=\bbl@dirlevel
912           \bbl@error{Multiple bidi settings inside a group}%
913           {I'll insert a new group, but expect wrong results.}%
914           \bgroup\aftergroup#2\aftergroup\egroup
915         \else
916           \ifcase\currentgroup\or % 0 bottom
917             \aftergroup#2% 1 simple {}
918           \or
919             \bgroup\aftergroup#2\aftergroup\egroup % 2 hbox
920           \or
921             \bgroup\aftergroup#2\aftergroup\egroup % 3 adj hbox
922           \or\or\or % vbox vtop align
923           \or
924             \bgroup\aftergroup#2\aftergroup\egroup % 7 noalign
925           \or\or\or\or\or\or % output math disc insert vcent mathchoice
926           \or
927             \aftergroup#2% 14 \begingroup
928           \else
929             \bgroup\aftergroup#2\aftergroup\egroup % 15 adj
930           \fi
931         \fi
932         \bbl@dirlevel\currentgrouplevel

```

```

933 \fi
934 #1%
935 \fi}
936 \def\bbl@pardir#1{\chardef\bbl@thepardir#1\relax}
937 \let\bbl@bodydir\@gobble
938 \let\bbl@pagedir\@gobble
939 \def\bbl@dirparastext{\chardef\bbl@thepardir\bbl@thetextdir}

```

The following command is executed only if there is a right-to-left script (once). It activates the `\everypar` hack for xetex, to properly handle the par direction. Note text and par dirs are decoupled to some extent (although not completely).

```

940 \def\bbl@xebidipar{%
941   \let\bbl@xebidipar\relax
942   \TeXeTstate\@ne
943   \def\bbl@xeverypar{%
944     \ifcase\bbl@thepardir
945       \ifcase\bbl@thetextdir\else\beginR\fi
946     \else
947       {\setbox\z@\lastbox\beginR\box\z@}%
948     \fi}%
949   \let\bbl@severypar\everypar
950   \newtoks\everypar
951   \everypar=\bbl@severypar
952   \bbl@severypar{\bbl@xeverypar\the\everypar}}
953 \ifnum\bbl@bidimode>200
954   \let\bbl@textdir\i\@gobbletwo
955   \let\bbl@xebidipar\@empty
956   \AddBabelHook{bidi}{foreign}{%
957     \def\bbl@tempa{\def\BabelText###1}%
958     \ifcase\bbl@thetextdir
959       \expandafter\bbl@tempa\expandafter{\BabelText\LR{##1}}}%
960     \else
961       \expandafter\bbl@tempa\expandafter{\BabelText\RL{##1}}}%
962     \fi}
963   \def\bbl@pardir#1{\ifcase#1\relax\setLR\else\setRL\fi}
964 \fi
965 \fi

```

A tool for weak L (mainly digits). We also disable warnings with `hyperref`.

```

966 \DeclareRobustCommand\babelsublr[1]{\leavevmode\bbl@textdir\z@#1}}
967 \AtBeginDocument{%
968   \ifx\pdfstringdefDisableCommands\@undefined\else
969     \ifx\pdfstringdefDisableCommands\relax\else
970       \pdfstringdefDisableCommands{\let\babelsublr\@firstofone}%
971     \fi
972   \fi}

```

7.10 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 `nor.sk.cfg` will be loaded when the language definition file `nor.sk.ldf` is loaded.

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

```

973 \bbl@trace{Local Language Configuration}
974 \ifx\loadlocalcfg\@undefined

```

```

975 \ifpackagewith{babel}{noconfigs}%
976 {\let\loadlocalcfg\@gobble}%
977 {\def\loadlocalcfg#1{%
978   \InputIfFileExists{#1.cfg}%
979   {\typeout{*****^J%
980             * Local config file #1.cfg used^^J%
981             *}}}%
982   \@empty}}
983 \fi

```

Just to be compatible with L^AT_EX 2.09 we add a few more lines of code. TODO. Necessary?
Correct place? Used by some ldf file?

```

984 \ifx\@unexpandable@protect\@undefined
985 \def\@unexpandable@protect{\noexpand\protect\noexpand}
986 \long\def\protected@write#1#2#3{%
987   \begingroup
988     \let\thepage\relax
989     #2%
990     \let\protect\@unexpandable@protect
991     \edef\reserved@a{\write#1{#3}}%
992     \reserved@a
993   \endgroup
994   \if@nobreak\ifvmode\nobreak\fi\fi}
995 \fi
996 %
997 % \subsection{Language options}
998 %
999 % Languages are loaded when processing the corresponding option
1000 % \textit{except} if a |main| language has been set. In such a
1001 % case, it is not loaded until all options has been processed.
1002 % The following macro inputs the ldf file and does some additional
1003 % checks (|\input| works, too, but possible errors are not caught).
1004 %
1005 % \begin{macrocode}
1006 \bbl@trace{Language options}
1007 \let\bbl@afterlang\relax
1008 \let\BabelModifiers\relax
1009 \let\bbl@loaded\@empty
1010 \def\bbl@load@language#1{%
1011   \InputIfFileExists{#1.ldf}%
1012   {\edef\bbl@loaded{\CurrentOption
1013     \ifx\bbl@loaded\@empty\else,\bbl@loaded\fi}%
1014     \expandafter\let\expandafter\bbl@afterlang
1015     \csname\CurrentOption.ldf-h@@k\endcsname
1016     \expandafter\let\expandafter\BabelModifiers
1017     \csname bbl@mod@\CurrentOption\endcsname}%
1018   {\bbl@error{%
1019     Unknown option '\CurrentOption'. Either you misspelled it\\%
1020     or the language definition file \CurrentOption.ldf was not found}}%
1021     Valid options are, among others: shorthands=, KeepShorthandsActive,\\%
1022     activeacute, activegrave, noconfigs, safe=, main=, math=\\%
1023     headfoot=, strings=, config=, hyphenmap=, or a language name.}}}

```

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

```

1024 \def\bbl@try@load@lang#1#2#3{%
1025   \IfFileExists{\CurrentOption.ldf}%

```



```

1026     {\bbl@load@language{\CurrentOption}}}%
1027     {\#1\bbl@load@language{\#2}\#3}}
1028 \DeclareOption{hebrew}{%
1029   \input{rlbabel.def}%
1030   \bbl@load@language{hebrew}}
1031 \DeclareOption{hungarian}{\bbl@try@load@lang{}{magyar}{}}
1032 \DeclareOption{lowersorbian}{\bbl@try@load@lang{}{lsorbian}{}}
1033 \DeclareOption{nynorsk}{\bbl@try@load@lang{}{norsk}{}}
1034 \DeclareOption{polutonikogreek}{%
1035   \bbl@try@load@lang{}{greek}{\languageattribute{greek}{polutoniko}}}%
1036 \DeclareOption{russian}{\bbl@try@load@lang{}{russianb}{}}
1037 \DeclareOption{ukrainian}{\bbl@try@load@lang{}{ukraineb}{}}
1038 \DeclareOption{uppersorbian}{\bbl@try@load@lang{}{usorbian}{}}

```

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

```

1039 \ifx\bbl@opt@config\@nnil
1040   \@ifpackagewith{babel}{noconfigs}{}%
1041   {\InputIfFileExists{bblopts.cfg}%
1042     {\typeout{*****^^J%
1043               * Local config file bblopts.cfg used^^J%
1044               *}}%
1045     {}}%
1046 \else
1047   \InputIfFileExists{\bbl@opt@config.cfg}%
1048   {\typeout{*****^^J%
1049             * Local config file \bbl@opt@config.cfg used^^J%
1050             *}}%
1051   {\bbl@error{%
1052     Local config file '\bbl@opt@config.cfg' not found}{%
1053     Perhaps you misspelled it.}}%
1054 \fi

```

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

```

1055 \let\bbl@tempc\relax
1056 \bbl@foreach\bbl@language@opts{%
1057   \ifcase\bbl@iniflag % Default
1058     \bbl@ifunset{ds@#1}%
1059     {\DeclareOption{#1}{\bbl@load@language{#1}}}%
1060     {}%
1061   \or % provide=*
1062     \@gobble % case 2 same as 1
1063   \or % provide+=*
1064     \bbl@ifunset{ds@#1}%
1065     {\IfFileExists{#1.ldf}{}%
1066      {\IfFileExists{babel-#1.tex}{\@namedef{ds@#1}}{}}}%
1067     {}%
1068   \bbl@ifunset{ds@#1}%
1069   {\def\bbl@tempc{#1}%
1070    \DeclareOption{#1}{%
1071      \ifnum\bbl@iniflag>\@ne

```

```

1072         \bbl@ldfinit
1073         \babelprovide[import]{#1}%
1074         \bbl@afterldf{}%
1075     \else
1076         \bbl@load@language{#1}%
1077     \fi}%
1078 {}%
1079 \or      % provide*=*
1080     \def\bbl@tempc{#1}%
1081     \bbl@ifunset{ds@#1}%
1082     {\DeclareOption{#1}{%
1083         \bbl@ldfinit
1084         \babelprovide[import]{#1}%
1085         \bbl@afterldf{}}}%
1086     {}%
1087 \fi}

```

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.

```

1088 \let\bbl@tempb\@nnil
1089 \bbl@foreach\@classoptionslist{%
1090     \bbl@ifunset{ds@#1}%
1091     {\IfFileExists{#1.ldf}{}%
1092      {\IfFileExists{babel-#1.tex}{\@namedef{ds@#1}{}}}%
1093     }%
1094     \bbl@ifunset{ds@#1}%
1095     {\def\bbl@tempb{#1}%
1096      \DeclareOption{#1}{%
1097          \ifnum\bbl@iniflag>\@ne
1098              \bbl@ldfinit
1099              \babelprovide[import]{#1}%
1100              \bbl@afterldf{}%
1101          \else
1102              \bbl@load@language{#1}%
1103          \fi}%
1104     }}

```

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

```

1105 \ifnum\bbl@iniflag=\z@ \else
1106     \ifx\bbl@opt@main\@nnil
1107         \ifx\bbl@tempc\relax
1108             \let\bbl@opt@main\bbl@tempb
1109         \else
1110             \let\bbl@opt@main\bbl@tempc
1111         \fi
1112     \fi
1113 \fi
1114 \ifx\bbl@opt@main\@nnil \else
1115     \expandafter
1116     \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
1117     \expandafter\let\csname ds@\bbl@opt@main\endcsname\@empty
1118 \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):

```

1119 \def\AfterBabelLanguage#1{%

```

```

1120 \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{}}
1121 \DeclareOption*{}
1122 \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.

```

1123 \bbl@trace{Option 'main'}
1124 \ifx\bbl@opt@main\nil
1125 \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
1126 \let\bbl@tempc\empty
1127 \bbl@for\bbl@tempb\bbl@tempa{%
1128   \bbl@xin{,\bbl@tempb,}{,\bbl@loaded,}%
1129   \ifin@\edef\bbl@tempc{\bbl@tempb}\fi}
1130 \def\bbl@tempa#1,#2\nil{\def\bbl@tempb{#1}}
1131 \expandafter\bbl@tempa\bbl@loaded,\nil
1132 \ifx\bbl@tempb\bbl@tempc\else
1133   \bbl@warning{%
1134     Last declared language option is '\bbl@tempc',\%
1135     but the last processed one was '\bbl@tempb'.\%
1136     The main language cannot be set as both a global\%
1137     and a package option. Use 'main=\bbl@tempc' as\%
1138     option. Reported}%
1139   \fi
1140 \else
1141   \ifodd\bbl@iniflag % case 1,3
1142     \bbl@ldfinit
1143     \let\CurrentOption\bbl@opt@main
1144     \bbl@exp{\bbl@babelprovide[import,main]{\bbl@opt@main}}
1145     \bbl@afterldf{}%
1146   \else % case 0,2
1147     \chardef\bbl@iniflag\z@ % Force ldf
1148     \expandafter\let\csname ds@\bbl@opt@main\endcsname\bbl@loadmain
1149     \ExecuteOptions{\bbl@opt@main}
1150     \DeclareOption*{}%
1151     \ProcessOptions*
1152   \fi
1153 \fi
1154 \def\AfterBabelLanguage{%
1155   \bbl@error
1156   {Too late for \string\AfterBabelLanguage}%
1157   {Languages have been loaded, so I can do nothing}}

```

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

```

1158 \ifx\bbl@main@language\undefined
1159   \bbl@info{%
1160     You haven't specified a language. I'll use 'nil'\%
1161     as the main language. Reported}
1162   \bbl@load@language{nil}
1163 \fi
1164 </package>
1165 <*core>

```

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

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

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

```
1166 \ifx\ldf@quit\@undefined\else
1167 \endinput\fi % Same line!
1168 <<Make sure ProvidesFile is defined>>
1169 \ProvidesFile{babel.def}[\<date>] \<version>] Babel common definitions]
```

The file babel.def expects some definitions made in the L^AT_EX 2_ε style file. So, In L^AT_EX 2.09 and Plain we must provide at least some predefined values as well some tools to set them (even if not all options are available). There are no package options, and therefore an alternative mechanism is provided. For the moment, only \babeloptionstrings and \babeloptionmath are provided, which can be defined before loading babel.

\BabelModifiers can be set too (but not sure it works).

```
1170 \ifx\AtBeginDocument\@undefined % TODO. change test.
1171 <<Emulate LaTeX>>
1172 \def\language#1{english}%
1173 \let\bbl@opt@shorthands\@nnil
1174 \def\bbl@ifshorthand#1#2#3{#2}%
1175 \let\bbl@language@opts\@empty
1176 \ifx\babeloptionstrings\@undefined
1177 \let\bbl@opt@strings\@nnil
1178 \else
1179 \let\bbl@opt@strings\babeloptionstrings
1180 \fi
1181 \def\BabelStringsDefault{generic}
1182 \def\bbl@tempa{normal}
1183 \ifx\babeloptionmath\bbl@tempa
1184 \def\bbl@mathnormal{\noexpand\textormath}
1185 \fi
1186 \def\AfterBabelLanguage#1#2{}
1187 \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
1188 \let\bbl@afterlang\relax
1189 \def\bbl@opt@safe{BR}
1190 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
1191 \ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
1192 \expandafter\newif\csname ifbbl@single\endcsname
1193 \chardef\bbl@bidimode\z@
1194 \fi
```

Exit immediately with 2.09. An error is raised by the sty file, but also try to minimize the number of errors.

```
1195 \ifx\bbl@trace\@undefined
1196 \let\LdfInit\endinput
1197 \def\ProvidesLanguage#1{\endinput}
1198 \endinput\fi % Same line!
```

And continue.

9 Multiple languages

This is not a separate file (switch.def) anymore.

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.

```
1199 <<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.

```
1200 \def\bbl@version{<<version>>}%
1201 \def\bbl@date{<<date>>}%
1202 \def\adddialect#1#2{%
1203   \global\chardef#1#2\relax
1204   \bbl@usehooks{adddialect}{#1}{#2}}%
1205 \begingroup
1206   \count@#1\relax
1207   \def\bbl@elt##1##2##3###4{%
1208     \ifnum\count@=##2\relax
1209       \bbl@info{\string#1 = using hyphenrules for ##1\\%
1210         (\string\language\the\count@)}%
1211       \def\bbl@elt####1####2####3####4{%
1212         \fi}%
1213       \bbl@cs{languages}%
1214     \endgroup}
```

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

```
1215 \def\bbl@fixname#1{%
1216   \begingroup
1217   \def\bbl@tempe{l@}%
1218   \edef\bbl@tempd{\noexpand\ifundefined{\noexpand\bbl@tempe#1}}%
1219   \bbl@tempd
1220     {\lowercase\expandafter{\bbl@tempd}%
1221     {\uppercase\expandafter{\bbl@tempd}%
1222     \@empty
1223     {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1224     {\uppercase\expandafter{\bbl@tempd}}}%
1225     {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1226     {\lowercase\expandafter{\bbl@tempd}}}%
1227     \@empty
1228     \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
1229   \bbl@tempd
1230   \bbl@exp{\bbl@usehooks{language}{\language}{#1}}}%
1231 \def\bbl@iflanguage#1{%
1232   \ifundefined{l@#1}{\@nolanerr{#1}\@gobble}\@firstofone}
```

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

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

```

1233 \def\bbl@bcpcase#1#2#3#4\@#5{%
1234   \ifx\@empty#3%
1235     \uppercase{\def#5{#1#2}}%
1236   \else
1237     \uppercase{\def#5{#1}}%
1238     \lowercase{\edef#5{#5#2#3#4}}%
1239   \fi}
1240 \def\bbl@bcplookup#1-#2-#3-#4\@{%
1241   \let\bbl@bcp\relax
1242   \lowercase{\def\bbl@tempa{#1}}%
1243   \ifx\@empty#2%
1244     \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1245   \else\ifx\@empty#3%
1246     \bbl@bcpcase#2\@empty\@empty\@{\bbl@tempb
1247     \IfFileExists{babel-\bbl@tempa-\bbl@tempb.ini}%
1248       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb}}%
1249     }%
1250     \ifx\bbl@bcp\relax
1251       \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1252     \fi
1253   \else
1254     \bbl@bcpcase#2\@empty\@empty\@{\bbl@tempb
1255     \bbl@bcpcase#3\@empty\@empty\@{\bbl@tempc
1256     \IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc.ini}%
1257       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb-\bbl@tempc}}%
1258     }%
1259     \ifx\bbl@bcp\relax
1260       \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1261       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
1262     }%
1263     \fi
1264     \ifx\bbl@bcp\relax
1265       \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1266       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
1267     }%
1268     \fi
1269     \ifx\bbl@bcp\relax
1270       \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1271     \fi
1272   \fi\fi}
1273 \let\bbl@initoload\relax
1274 \def\bbl@provide@locale{%
1275   \ifx\babelprovide\undefined
1276     \bbl@error{For a language to be defined on the fly 'base'\\%
1277       is not enough, and the whole package must be\\%
1278       loaded. Either delete the 'base' option or\\%
1279       request the languages explicitly}%
1280     {See the manual for further details.}%
1281   \fi
1282 % TODO. Option to search if loaded, with \LocaleForEach
1283 \let\bbl@auxname\language\name % Still necessary. TODO
1284 \bbl@ifunset{\bbl@bcp@map@\language\name}{}% Move uplevel??
1285 {\edef\language\name{\@nameuse{\bbl@bcp@map@\language\name}}}%
1286 \ifbbl@bcpallowed

```

```

1287 \expandafter\ifx\csname date\language\endcsname\relax
1288 \expandafter
1289 \bbl@bcplookup\language-\@empty-\@empty-\@empty\@
1290 \ifx\bbl@bcp\relax\else % Returned by \bbl@bcplookup
1291 \edef\language{\bbl@bcp@prefix\bbl@bcp}%
1292 \edef\localename{\bbl@bcp@prefix\bbl@bcp}%
1293 \expandafter\ifx\csname date\language\endcsname\relax
1294 \let\bbl@initoload\bbl@bcp
1295 \bbl@exp{\bbl@babelprovide[\bbl@autoload@bcptoptions]{\language}}%
1296 \let\bbl@initoload\relax
1297 \fi
1298 \bbl@csarg\xdef{bcp@map@\bbl@bcp}{\localename}%
1299 \fi
1300 \fi
1301 \fi
1302 \expandafter\ifx\csname date\language\endcsname\relax
1303 \IfFileExists{babel-\language.tex}%
1304 {\bbl@exp{\bbl@babelprovide[\bbl@autoload@options]{\language}}}%
1305 {}%
1306 \fi}

```

`\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.

```

1307 \def\iflanguage#1{%
1308 \bbl@iflanguage{#1}{%
1309 \ifnum\csname l@#1\endcsname=\language
1310 \expandafter\@firstoftwo
1311 \else
1312 \expandafter\@secondoftwo
1313 \fi}}

```

9.1 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.

```

1314 \let\bbl@select@type\z@
1315 \edef\selectlanguage{%
1316 \noexpand\protect
1317 \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`.

```

1318 \ifx\@undefined\protect\let\protect\relax\fi

```

The following definition is preserved for backwards compatibility. It is related to a trick for 2.09.

```

1319 \let\xstring\string

```

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

`\bbl@pop@language` But when the language change happens *inside* a group the end of the group doesn't write anything to the auxiliary files. Therefore we need TeX's `aftergroup` mechanism to help us. The command `\aftergroup` stores the token immediately following it to be executed when the current group is closed. So we define a temporary control sequence `\bbl@pop@language` to be executed at the end of the group. It calls `\bbl@set@language` with the name of the current language as its argument.

`\bbl@language@stack` The previous solution works for one level of nesting groups, but as soon as more levels are used it is no longer adequate. For that case we need to keep track of the nested languages using a stack mechanism. This stack is called `\bbl@language@stack` and initially empty.

```
1320 \def\bbl@language@stack{}
```

When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.

`\bbl@push@language` The stack is simply a list of languagenames, separated with a '+' sign; the push function can
`\bbl@pop@language` be simple:

```
1321 \def\bbl@push@language{%
1322   \ifx\language\undefined\else
1323     \xdef\bbl@language@stack{\language+\bbl@language@stack}%
1324   \fi}
```

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 in `\bbl@language@stack`.

```
1325 \def\bbl@pop@lang#1+#2\@@{%
1326   \edef\language{#1}%
1327   \xdef\bbl@language@stack{#2}}
```

The reason for the somewhat weird arrangement of arguments to the helper function is the fact it is called in the following way. This means that before `\bbl@pop@lang` is executed TeX first *expands* the stack, stored in `\bbl@language@stack`. The result of that is that the argument string of `\bbl@pop@lang` contains one or more language names, each followed by a '+'-sign (zero language names won't occur as this macro will only be called after something has been pushed on the stack).

```
1328 \let\bbl@ifrestoring\@secondoftwo
1329 \def\bbl@pop@language{%
1330   \expandafter\bbl@pop@lang\bbl@language@stack\@@
1331   \let\bbl@ifrestoring\@firstoftwo
1332   \expandafter\bbl@set@language\expandafter{\language}%
1333   \let\bbl@ifrestoring\@secondoftwo}
```

Once the name of the previous language is retrieved from the stack, it is fed to `\bbl@set@language` to do the actual work of switching everything that needs switching.

An alternative way to identify languages (in the babel sense) with a numerical value is introduced in 3.30. This is one of the first steps for a new interface based on the concept of locale, which explains the name of `\localeid`. This means `\l@...` will be reserved for hyphenation patterns (so that two locales can share the same rules).

```
1334 \chardef\localeid\z@
1335 \def\bbl@id@last{0} % No real need for a new counter
1336 \def\bbl@id@assign{%
1337   \bbl@ifunset\bbl@id@\language}%
1338   {\count\bbl@id@last\relax
```



```

1339 \advance\count@ \@ne
1340 \bbl@csarg\chardef{id@@\language}\count@
1341 \edef\bbl@id@last{\the\count@}%
1342 \ifcase\bbl@engine\or
1343 \directlua{
1344     Babel = Babel or {}
1345     Babel.locale_props = Babel.locale_props or {}
1346     Babel.locale_props[\bbl@id@last] = {}
1347     Babel.locale_props[\bbl@id@last].name = '\language'
1348 }%
1349 \fi}%
1350 {}%
1351 \chardef\localeid\bbl@c1{id@}}

```

The unprotected part of `\selectlanguage`.

```

1352 \expandafter\def\csname selectlanguage \endcsname#1{%
1353 \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\tw@\fi
1354 \bbl@push@language
1355 \aftergroup\bbl@pop@language
1356 \bbl@set@language{#1}}

```

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

```

1357 \def\BabelContentsFiles{toc,lof,lot}
1358 \def\bbl@set@language#1{% from selectlanguage, pop@
1359 % The old buggy way. Preserved for compatibility.
1360 \edef\language{%
1361 \ifnum\escapechar=\expandafter`\string#1\@empty
1362 \else\string#1\@empty\fi}%
1363 \ifcat\relax\noexpand#1%
1364 \expandafter\ifx\csname date\language\endcsname\relax
1365 \edef\language{#1}%
1366 \let\localename\language
1367 \else
1368 \bbl@info{Using '\string\language' instead of 'language' is\\%
1369 deprecated. If what you want is to use a\\%
1370 macro containing the actual locale, make\\%
1371 sure it does not not match any language.\\%
1372 Reported}%
1373 % I'll\\%
1374 % try to fix '\string\localename', but I cannot promise\\%
1375 % anything. Reported}%
1376 \ifx\scantokens\undefined
1377 \def\localename{??}%
1378 \else
1379 \scantokens\expandafter{\expandafter
1380 \def\expandafter\localename\expandafter{\language}}%
1381 \fi
1382 \fi
1383 \else
1384 \def\localename{#1}% This one has the correct catcodes
1385 \fi

```

```

1386 \select@language{\language}%
1387 % write to auxs
1388 \expandafter\ifx\csname date\language\endcsname\relax\else
1389   \if@filesw
1390     \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
1391       \protected@write\@auxout{}\string\babel@aux{\bbl@auxname}{}}%
1392     \fi
1393     \bbl@usehooks{write}}}%
1394   \fi
1395 \fi}
1396 %
1397 \newif\ifbbl@bcpallowed
1398 \bbl@bcpallowedfalse
1399 \def\select@language#1{% from set@, babel@aux
1400   % set hmap
1401   \ifnum\bbl@hymapsel=\@cclv\chardef\bbl@hymapsel4\relax\fi
1402   % set name
1403   \edef\language{#1}%
1404   \bbl@fixname\language
1405   % TODO. name@map must be here?
1406   \bbl@provide@locale
1407   \bbl@iflanguage\language{%
1408     \expandafter\ifx\csname date\language\endcsname\relax
1409       \bbl@error
1410       {Unknown language '\language'. Either you have\\%
1411        misspelled its name, it has not been installed,\\%
1412        or you requested it in a previous run. Fix its name,\\%
1413        install it or just rerun the file, respectively. In\\%
1414        some cases, you may need to remove the aux file}%
1415       {You may proceed, but expect wrong results}%
1416     \else
1417       % set type
1418       \let\bbl@select@type\z@
1419       \expandafter\bbl@switch\expandafter{\language}%
1420     \fi}}
1421 \def\babel@aux#1#2{% TODO. See how to avoid undefined nil's
1422   \select@language{#1}%
1423   \bbl@foreach\BabelContentsFiles{%
1424     \@writefile{##1}{\babel@toc{#1}{#2}}}% % TODO - ok in plain?
1425 \def\babel@toc#1#2{%
1426   \select@language{#1}}

```

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.

```

1427 \newif\ifbbl@usedategroup

```

```

1428 \def\bbl@switch#1{% from select@, foreign@
1429 % make sure there is info for the language if so requested
1430 \bbl@ensureinfo{#1}%
1431 % restore
1432 \originalTeX
1433 \expandafter\def\expandafter\originalTeX\expandafter{%
1434   \csname noextras#1\endcsname
1435   \let\originalTeX\@empty
1436   \babel@beginsave}%
1437 \bbl@usehooks{afterreset}}}%
1438 \languageshorthands{none}%
1439 % set the locale id
1440 \bbl@id@assign
1441 % switch captions, date
1442 % No text is supposed to be added here, so we remove any
1443 % spurious spaces.
1444 \bbl@bsphack
1445   \ifcase\bbl@select@type
1446     \csname captions#1\endcsname\relax
1447     \csname date#1\endcsname\relax
1448   \else
1449     \bbl@xin@{,captions,}{, \bbl@select@opts,}%
1450     \ifin@
1451       \csname captions#1\endcsname\relax
1452     \fi
1453     \bbl@xin@{,date,}{, \bbl@select@opts,}%
1454     \ifin@ % if \foreign... within \<lang>date
1455       \csname date#1\endcsname\relax
1456     \fi
1457   \fi
1458 \bbl@esphack
1459 % switch extras
1460 \bbl@usehooks{beforeextras}}}%
1461 \csname extras#1\endcsname\relax
1462 \bbl@usehooks{afterextras}}}%
1463 % > babel-ensure
1464 % > babel-sh-<short>
1465 % > babel-bidi
1466 % > babel-fontspec
1467 % hyphenation - case mapping
1468 \ifcase\bbl@opt@hyphenmap\or
1469   \def\BabelLower##1##2{\lccode##1=##2\relax}%
1470   \ifnum\bbl@hymapsel>4\else
1471     \csname\language @bbl@hyphenmap\endcsname
1472   \fi
1473   \chardef\bbl@opt@hyphenmap\z@
1474 \else
1475   \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
1476     \csname\language @bbl@hyphenmap\endcsname
1477   \fi
1478 \fi
1479 \let\bbl@hymapsel\@cclv
1480 % hyphenation - select patterns
1481 \bbl@patterns{#1}%
1482 % hyphenation - allow stretching with babelnohyphens
1483 \ifnum\language=\l@babelnohyphens
1484   \babel@savevariable\emergencystretch
1485   \emergencystretch\maxdimen
1486   \babel@savevariable\hbadness

```

```

1487 \hbadness\@M
1488 \fi
1489 % hyphenation - mins
1490 \babel@savevariable\lefthyphenmin
1491 \babel@savevariable\righthyphenmin
1492 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1493 \set@hyphenmins\tw@\thr@\relax
1494 \else
1495 \expandafter\expandafter\expandafter\set@hyphenmins
1496 \csname #1hyphenmins\endcsname\relax
1497 \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.

```

1498 \long\def\otherlanguage#1{%
1499 \ifnum\bbbl@hymapsel=\@cclv\let\bbbl@hymapsel\thr@\fi
1500 \csname selectlanguage \endcsname{#1}%
1501 \ignorespaces}

```

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

```

1502 \long\def\endotherlanguage{%
1503 \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`.

```

1504 \expandafter\def\csname otherlanguage*\endcsname{%
1505 \@ifnextchar[\bbbl@otherlanguage@s{\bbbl@otherlanguage@s[]}}
1506 \def\bbbl@otherlanguage@s[#1]#2{%
1507 \ifnum\bbbl@hymapsel=\@cclv\chardef\bbbl@hymapsel4\relax\fi
1508 \def\bbbl@select@opts{#1}%
1509 \foreign@language{#2}}

```

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

```

1510 \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`. `\bbbl@beforeforeign` is a trick to fix a bug in bidi texts. `\foreignlanguage` is supposed to be a ‘text’ command, and therefore it must emit a `\leavevmode`, but it does not, and therefore the indent is placed on the opposite margin. For backward compatibility, however, it is done only if a right-to-left script is requested; otherwise, it is no-op. (3.11) `\foreignlanguage*` is a temporary, experimental macro for a few lines with a different script direction, while preserving the paragraph format (thank the braces around

\par, things like \hangindent are not reset). Do not use it in production, because its semantics and its syntax may change (and very likely will, or even it could be removed altogether). Currently it enters in vmode and then selects the language (which in turn sets the paragraph direction).

(3.11) Also experimental are the hook foreign and foreign*. With them you can redefine \BabelText which by default does nothing. Its behavior is not well defined yet. So, use it in horizontal mode only if you do not want surprises.

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

```

1511 \providecommand\bbl@beforeforeign{
1512 \edef\foreignlanguage{%
1513   \noexpand\protect
1514   \expandafter\noexpand\csname foreignlanguage \endcsname}
1515 \expandafter\def\csname foreignlanguage \endcsname{%
1516   \@ifstar\bbl@foreign@s\bbl@foreign@x}
1517 \providecommand\bbl@foreign@x[3][{}]{%
1518   \begingroup
1519     \def\bbl@select@opts{#1}%
1520     \let\BabelText\@firstofone
1521     \bbl@beforeforeign
1522     \foreign@language{#2}%
1523     \bbl@usehooks{foreign}{}%
1524     \BabelText{#3}% Now in horizontal mode!
1525   \endgroup}
1526 \def\bbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@@par
1527   \begingroup
1528     {\par}%
1529     \let\BabelText\@firstofone
1530     \foreign@language{#1}%
1531     \bbl@usehooks{foreign*}{}%
1532     \bbl@dirparastext
1533     \BabelText{#2}% Still in vertical mode!
1534     {\par}%
1535   \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.

```

1536 \def\foreign@language#1{%
1537   % set name
1538   \edef\language{#1}%
1539   \ifbbl@usedategroup
1540     \bbl@add\bbl@select@opts{,date,}%
1541     \bbl@usedategroupfalse
1542   \fi
1543   \bbl@fixname\language
1544   % TODO. name@map here?
1545   \bbl@provide@locale
1546   \bbl@iflanguage\language{%
1547     \expandafter\ifx\csname date\language\endcsname\relax
1548       \bbl@warning % TODO - why a warning, not an error?
1549       {Unknown language `#1'. Either you have\\%
1550        misspelled its name, it has not been installed,\\%
1551        or you requested it in a previous run. Fix its name,\\%
1552        install it or just rerun the file, respectively. In\\%
1553        some cases, you may need to remove the aux file.\\%
1554        I'll proceed, but expect wrong results.\\%

```

```

1555         Reported}%
1556     \fi
1557     % set type
1558     \let\bbl@select@type\@ne
1559     \expandafter\bbl@switch\expandafter{\language}%

```

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

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

```

1560 \let\bbl@hyphlist\@empty
1561 \let\bbl@hyphenation@relax
1562 \let\bbl@pttnlist\@empty
1563 \let\bbl@patterns@relax
1564 \let\bbl@hymapsel=\@ccclv
1565 \def\bbl@patterns#1{%
1566     \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
1567         \csname l@#1\endcsname
1568         \edef\bbl@tempa{#1}%
1569     \else
1570         \csname l@#1:\f@encoding\endcsname
1571         \edef\bbl@tempa{#1:\f@encoding}%
1572     \fi
1573     \@expandtwoargs\bbl@usehooks{patterns}{#1}{\bbl@tempa}%
1574     % > luatex
1575     \@ifundefined{bbl@hyphenation@}{% Can be \relax!
1576         \begingroup
1577         \bbl@xin@{\number\language,}{\bbl@hyphlist}%
1578         \ifin@else
1579             \@expandtwoargs\bbl@usehooks{hyphenation}{#1}{\bbl@tempa}%
1580             \hyphenation{%
1581                 \bbl@hyphenation@
1582                 \@ifundefined{bbl@hyphenation@#1}%
1583                 \@empty
1584                 {\space\csname bbl@hyphenation@#1\endcsname}}%
1585             \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
1586         \fi
1587     \endgroup}}

```

hyphenrules The environment `hyphenrules` can be used to select *just* the hyphenation rules. This environment does *not* change `\language` and when the hyphenation rules specified were not loaded it has no effect. Note however, `\lccode's` and font encodings are not set at all, so in most cases you should use other language*.

```

1588 \def\hyphenrules#1{%
1589     \edef\bbl@tempf{#1}%
1590     \bbl@fixname\bbl@tempf
1591     \bbl@iflanguage\bbl@tempf{%
1592         \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
1593         \ifx\languageshorthands\undefined\else
1594             \languageshorthands{none}%
1595         \fi
1596         \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax

```

```

1597     \set@hyphenmins\tw@\thr@@\relax
1598   \else
1599     \expandafter\expandafter\expandafter\set@hyphenmins
1600     \csname\bbl@tempf hyphenmins\endcsname\relax
1601   \fi}}
1602 \let\endhyphenrules\@empty

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

1603 \def\providehyphenmins#1#2{%
1604   \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1605     \@namedef{#1hyphenmins}{#2}%
1606   \fi}

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

1607 \def\set@hyphenmins#1#2{%
1608   \lefthyphenmin#1\relax
1609   \righthyphenmin#2\relax}

\ProvidesLanguage  The identification code for each file is something that was introduced in LATEX 2ε. When the
                    command \ProvidesFile does not exist, a dummy definition is provided temporarily. For
                    use in the language definition file the command \ProvidesLanguage is defined by babel.
                    Depending on the format, ie, on if the former is defined, we use a similar definition or not.

1610 \ifx\ProvidesFile\@undefined
1611   \def\ProvidesLanguage#1[#2 #3 #4]{%
1612     \wlog{Language: #1 #4 #3 <#2>}%
1613   }
1614 \else
1615   \def\ProvidesLanguage#1{%
1616     \begingroup
1617       \catcode`\ 10 %
1618       \@makeother\/%
1619       \@ifnextchar[%]
1620         {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}
1621   \def\@provideslanguage#1[#2]{%
1622     \wlog{Language: #1 #2}%
1623     \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
1624   \endgroup}
1625 \fi

\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.

1626 \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 initializes the save mechanism, \babel@beginsave, is not considered to be
undefined.

1627 \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’:

1628 \providecommand\setlocale{%
1629   \bbl@error
1630   {Not yet available}%

```

```

1631 {Find an armchair, sit down and wait}}
1632 \let\uselocale\setlocale
1633 \let\locale\setlocale
1634 \let\selectlocale\setlocale
1635 \let\localename\setlocale
1636 \let\textlocale\setlocale
1637 \let\textlanguage\setlocale
1638 \let\language\setlocale

```

9.2 Errors

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

`\@noopterr` When the package was loaded without options not everything will work as expected. An error message is issued in that case.
 When the format knows about `\PackageError` it must be $\LaTeX 2_{\epsilon}$, so we can safely use its error handling interface. Otherwise we'll have to 'keep it simple'.
 Infos are not written to the console, but on the other hand many people think warnings are errors, so a further message type is defined: an important info which is sent to the console.

```

1639 \edef\bbl@nulllanguage{\string\language=0}
1640 \ifx\PackageError\undefined % TODO. Move to Plain
1641 \def\bbl@error#1#2{%
1642 \begingroup
1643 \newlinechar=`^^J
1644 \def\{^^J(babel) }%
1645 \errhelp{#2}\errmessage{\{#1}%
1646 \endgroup}
1647 \def\bbl@warning#1{%
1648 \begingroup
1649 \newlinechar=`^^J
1650 \def\{^^J(babel) }%
1651 \message{\{#1}%
1652 \endgroup}
1653 \let\bbl@infowarn\bbl@warning
1654 \def\bbl@info#1{%
1655 \begingroup
1656 \newlinechar=`^^J
1657 \def\{^^J}%
1658 \wlog{#1}%
1659 \endgroup}
1660 \fi
1661 \def\bbl@nocaption{\protect\bbl@nocaption@i}
1662 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
1663 \global\@namedef{#2}{\textbf{?#1?}}%
1664 \@nameuse{#2}%
1665 \bbl@warning{%
1666 \@backslashchar#2 not set. Please, define it\\%
1667 after the language has been loaded (typically\\%
1668 in the preamble) with something like:\\%
1669 \string\renewcommand\@backslashchar#2{..}\\%
1670 Reported}}
1671 \def\bbl@tentative{\protect\bbl@tentative@i}
1672 \def\bbl@tentative@i#1{%

```



```

1673 \bbl@warning{%
1674   Some functions for '#1' are tentative.\\%
1675   They might not work as expected and their behavior\\%
1676   could change in the future.\\%
1677   Reported}}
1678 \def\nolanerr#1{%
1679   \bbl@error
1680   {You haven't defined the language #1\space yet.\\%
1681     Perhaps you misspelled it or your installation\\%
1682     is not complete}%
1683   {Your command will be ignored, type <return> to proceed}}
1684 \def\nopatterns#1{%
1685   \bbl@warning
1686   {No hyphenation patterns were preloaded for\\%
1687     the language '#1' into the format.\\%
1688     Please, configure your TeX system to add them and\\%
1689     rebuild the format. Now I will use the patterns\\%
1690     preloaded for \bbl@nulllanguage\space instead}}
1691 \let\bbl@usehooks\@gobbletwo
1692 \ifx\bbl@onlyswitch\@empty\endinput\fi
1693 % Here ended switch.def

Here ended switch.def.

1694 \ifx\directlua\@undefined\else
1695   \ifx\bbl@luapatterns\@undefined
1696     \input luababel.def
1697   \fi
1698 \fi
1699 <<Basic macros>>
1700 \bbl@trace{Compatibility with language.def}
1701 \ifx\bbl@languages\@undefined
1702   \ifx\directlua\@undefined
1703     \openin1 = language.def % TODO. Remove hardcoded number
1704     \ifeof1
1705       \closein1
1706       \message{I couldn't find the file language.def}
1707     \else
1708       \closein1
1709       \begingroup
1710         \def\addlanguage#1#2#3#4#5{%
1711           \expandafter\ifx\csname lang@#1\endcsname\relax\else
1712             \global\expandafter\let\csname l@#1\expandafter\endcsname
1713             \csname lang@#1\endcsname
1714           \fi}%
1715         \def\uselanguage#1{}%
1716         \input language.def
1717       \endgroup
1718     \fi
1719   \fi
1720   \chardef\l@english\z@
1721 \fi

```

\addto 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. Note there is an inconsistency, because the assignment in the last branch is global.

```

1722 \def\addto#1#2{%
1723   \ifx#1\undefined
1724     \def#1{#2}%
1725   \else
1726     \ifx#1\relax
1727       \def#1{#2}%
1728     \else
1729       {\toks@\expandafter{#1#2}%
1730        \xdef#1{\the\toks@}}%
1731   \fi
1732 \fi}

```

The macro `\initiate@active@char` below takes all the necessary actions to make its argument a shorthand character. The real work is performed once for each character. But first we define a little tool. `TODO`. Always used with additional expansions. Move them here? Move the macro to basic?

```

1733 \def\bbl@withactive#1#2{%
1734   \begingroup
1735   \lccode`~=#2\relax
1736   \lowercase{\endgroup#1~}}

```

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

```

1737 \def\bbl@redefine#1{%
1738   \edef\bbl@tempa{\bbl@stripslash#1}%
1739   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1740   \expandafter\def\csname\bbl@tempa\endcsname{
1741   \@onlypreamble\bbl@redefine

```

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

```

1742 \def\bbl@redefine@long#1{%
1743   \edef\bbl@tempa{\bbl@stripslash#1}%
1744   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1745   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname{
1746   \@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_`.

```

1747 \def\bbl@redefineroobust#1{%
1748   \edef\bbl@tempa{\bbl@stripslash#1}%
1749   \bbl@ifunset{\bbl@tempa\space}%
1750   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
1751    \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
1752   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
1753   \@namedef{\bbl@tempa\space}}
1754 \@onlypreamble\bbl@redefineroobust

```

9.3 Hooks

Admittedly, the current implementation is a somewhat simplistic and does very little to catch errors, but it is meant for developers, after all. `\bbl@usehooks` is the commands used

by babel to execute hooks defined for an event.

```

1755 \bbl@trace{Hooks}
1756 \newcommand\AddBabelHook[3][{}]{%
1757   \bbl@ifunset{\bbl@hk@#2}{\EnableBabelHook{#2}}{}%
1758   \def\bbl@tempa##1,#3=##2,##3\@empty{\def\bbl@tempb{##2}}%
1759   \expandafter\bbl@tempa\bbl@evargs,#3=,\@empty
1760   \bbl@ifunset{\bbl@ev@#2@#3@#1}%
1761     {\bbl@csarg\bbl@add{ev@#3@#1}{\bbl@elth{#2}}}%
1762     {\bbl@csarg\let{ev@#2@#3@#1}\relax}%
1763   \bbl@csarg\newcommand{ev@#2@#3@#1}[\bbl@tempb]}
1764 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
1765 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
1766 \def\bbl@usehooks#1#2{%
1767   \def\bbl@elth##1{%
1768     \bbl@cs{hk@##1}{\bbl@cs{ev@##1@#1@#2}}%
1769     \bbl@cs{ev@#1@}%
1770     \ifx\language\@undefined\else % Test required for Plain (?)
1771       \def\bbl@elth##1{%
1772         \bbl@cs{hk@##1}{\bbl@cl{ev@##1@#1@#2}}%
1773         \bbl@cl{ev@#1}%
1774       \fi}

```

To ensure forward compatibility, arguments in hooks are set implicitly. So, if a further argument is added in the future, there is no need to change the existing code. Note events intended for hyphen.cfg are also loaded (just in case you need them for some reason).

```

1775 \def\bbl@evargs{% <- don't delete this comma
1776   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
1777   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
1778   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
1779   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
1780   beforestart=0,language=2}

```

`\babelensure` The user command just parses the optional argument and creates a new macro named `\bbl@e@<language>`. We register a hook at the `afterextras` event which just executes this macro in a “complete” selection (which, if undefined, is `\relax` and does nothing). This part is somewhat involved because we have to make sure things are expanded the correct number of times.

The macro `\bbl@e@<language>` contains `\bbl@ensure{\include}{\exclude}{\fontenc}`, which in turn loops over the macros names in `\bbl@captionslist`, excluding (with the help of `\in@`) those in the exclude list. If the `fontenc` is given (and not `\relax`), the `\fontencoding` is also added. Then we loop over the include list, but if the macro already contains `\foreignlanguage`, nothing is done. Note this macro (1) is not restricted to the preamble, and (2) changes are local.

```

1781 \bbl@trace{Defining babelensure}
1782 \newcommand\babelensure[2][{}]{% TODO - revise test files
1783   \AddBabelHook{babel-ensure}{afterextras}{%
1784     \ifcase\bbl@select@type
1785       \bbl@cl{e}%
1786     \fi}%
1787   \begingroup
1788     \let\bbl@ens@include\@empty
1789     \let\bbl@ens@exclude\@empty
1790     \def\bbl@ens@fontenc{\relax}%
1791     \def\bbl@tempb##1{%
1792       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
1793     \edef\bbl@tempa{\bbl@tempb#1\@empty}%
1794     \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ens@##1}{##2}}%

```

```

1795 \bbl@foreach\bbl@tempa{\bbl@tempb##1\@@}%
1796 \def\bbl@tempc{\bbl@ensure}%
1797 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1798 \expandafter{\bbl@ens@include}}%
1799 \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1800 \expandafter{\bbl@ens@exclude}}%
1801 \toks@\expandafter{\bbl@tempc}%
1802 \bbl@exp{%
1803 \endgroup
1804 \def<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
1805 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
1806 \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
1807 \ifx##1\undefined % 3.32 - Don't assume the macro exists
1808 \edef##1{\noexpand\bbl@nocaption
1809 {\bbl@stripslash##1}{\language\language\bbl@stripslash##1}}%
1810 \fi
1811 \ifx##1\@empty\else
1812 \in@{##1}{#2}%
1813 \ifin@ \else
1814 \bbl@ifunset{\bbl@ensure@\language}%
1815 {\bbl@exp{%
1816 \\\DeclareRobustCommand<bbl@ensure@\language>[1]{%
1817 \\\foreignlanguage{\language}%
1818 {\ifx\relax#3\else
1819 \\\fontencoding{#3}\selectfont
1820 \fi
1821 #####1}}}%
1822 {}}%
1823 \toks@\expandafter{##1}%
1824 \edef##1{%
1825 \bbl@csarg\noexpand{ensure@\language}%
1826 {\the\toks@}}%
1827 \fi
1828 \expandafter\bbl@tempb
1829 \fi}%
1830 \expandafter\bbl@tempb\bbl@captionslist\today\@empty
1831 \def\bbl@tempa##1{% elt for include list
1832 \ifx##1\@empty\else
1833 \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
1834 \ifin@ \else
1835 \bbl@tempb##1\@empty
1836 \fi
1837 \expandafter\bbl@tempa
1838 \fi}%
1839 \bbl@tempa#1\@empty}
1840 \def\bbl@captionslist{%
1841 \prefacename\refname\abstractname\bibname\chaptername\appendixname
1842 \contentsname\listfigurename\listtablename\indexname\figurename
1843 \tablename\partname\encname\ccname\headtoname\pagename\seename
1844 \alsoname\proofname\glossaryname}

```

9.4 Setting up language files

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

the at-sign. We make sure that it is a ‘letter’ during the processing of the file. We also save its name as the last called option, even if not loaded.

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

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

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

Finally we check `\originalTeX`.

```

1845 \bbl@trace{Macros for setting language files up}
1846 \def\bbl@ldfinit{%
1847   \let\bbl@screset\@empty
1848   \let\BabelStrings\bbl@opt@string
1849   \let\BabelOptions\@empty
1850   \let\BabelLanguages\relax
1851   \ifx\originalTeX\@undefined
1852     \let\originalTeX\@empty
1853   \else
1854     \originalTeX
1855   \fi}
1856 \def\LdfInit#1#2{%
1857   \chardef\atcatcode=\catcode`\@
1858   \catcode`\@=11\relax
1859   \chardef\eqcatcode=\catcode`\=
1860   \catcode`\==12\relax
1861   \expandafter\if\expandafter\@backslashchar
1862     \expandafter\@car\string#2\@nil
1863     \ifx#2\@undefined\else
1864       \ldf@quit{#1}%
1865     \fi
1866   \else
1867     \expandafter\ifx\csname#2\endcsname\relax\else
1868       \ldf@quit{#1}%
1869     \fi
1870   \fi
1871   \bbl@ldfinit}

```

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

```

1872 \def\ldf@quit#1{%
1873   \expandafter\main@language\expandafter{#1}%
1874   \catcode`\@=\atcatcode \let\atcatcode\relax
1875   \catcode`\==\eqcatcode \let\eqcatcode\relax
1876   \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.

```

1877 \def\bbl@afterldf#1{% TODO. Merge into the next macro? Unused elsewhere
1878   \bbl@afterlang

```

```

1879 \let\bbl@afterlang\relax
1880 \let\BabelModifiers\relax
1881 \let\bbl@screset\relax}%
1882 \def\ldf@finish#1{%
1883 \ifx\loadlocalcfg\undefined\else % For LaTeX 209
1884 \loadlocalcfg{#1}%
1885 \fi
1886 \bbl@afterldf{#1}%
1887 \expandafter\main@language\expandafter{#1}%
1888 \catcode`\@=\atcatcode \let\atcatcode\relax
1889 \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`.

```

1890 \onlypreamble\LdfInit
1891 \onlypreamble\ldf@quit
1892 \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.

```

1893 \def\main@language#1{%
1894 \def\bbl@main@language{#1}%
1895 \let\language\main@language % TODO. Set localename
1896 \bbl@id@assign
1897 \bbl@patterns{\language}}

```

We also have to make sure that some code gets executed at the beginning of the document, either when the aux file is read or, if it does not exist, when the `\AtBeginDocument` is executed. Languages do not set `\pagedir`, so we set here for the whole document to the main `\bodydir`.

```

1898 \def\bbl@beforestart{%
1899 \bbl@usehooks{beforestart}}}%
1900 \global\let\bbl@beforestart\relax}
1901 \AtBeginDocument{%
1902 \@nameuse{bbl@beforestart}%
1903 \if@filesw
1904 \providecommand\babel@aux[2]{}%
1905 \immediate\write\@mainaux{%
1906 \string\providecommand\string\babel@aux[2]{}%
1907 \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}}%
1908 \fi
1909 \expandafter\selectlanguage\expandafter{\bbl@main@language}%
1910 \ifbbl@single % must go after the line above.
1911 \renewcommand\selectlanguage[1]{}%
1912 \renewcommand\foreignlanguage[2]{#2}%
1913 \global\let\babel@aux\@gobbletwo % Also as flag
1914 \fi
1915 \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

1916 \def\select@language@x#1{%
1917 \ifcase\bbl@select@type
1918 \bbl@ifsamestring\language{#1}{\select@language{#1}}%
1919 \else
1920 \select@language{#1}%
1921 \fi}

```

9.5 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 \TeX 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.

```
1922 \bbl@trace{Shorhands}
1923 \def\bbl@add@special#1{% 1:a macro like "\", \?, etc.
1924   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
1925   \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
1926   \ifx\nfss@catcodes\undefined\else % TODO - same for above
1927     \begingroup
1928       \catcode`#1\active
1929       \nfss@catcodes
1930       \ifnum\catcode`#1=\active
1931         \endgroup
1932         \bbl@add\nfss@catcodes{\@makeother#1}%
1933       \else
1934         \endgroup
1935       \fi
1936   \fi}
```

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

```
1937 \def\bbl@remove@special#1{%
1938   \begingroup
1939   \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
1940     \else\noexpand##1\noexpand##2\fi}%
1941   \def\do{\x\do}%
1942   \def\@makeother{\x\@makeother}%
1943   \edef\x{\endgroup
1944     \def\noexpand\dospecials{\dospecials}%
1945     \expandafter\ifx\csname @sanitize\endcsname\relax\else
1946       \def\noexpand\@sanitize{\@sanitize}%
1947     \fi}%
1948   \x}
```

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

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

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

```

1949 \def\bbl@active@def#1#2#3#4{%
1950   \namedef{#3#1}{%
1951     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
1952       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
1953     \else
1954       \bbl@afterfi\csname#2@sh@#1\endcsname
1955     \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.

```

1956   \long\namedef{#3@arg#1}##1{%
1957     \expandafter\ifx\csname#2@sh@#1\string##1\endcsname\relax
1958       \bbl@afterelse\csname#4#1\endcsname##1%
1959     \else
1960       \bbl@afterfi\csname#2@sh@#1\string##1\endcsname
1961     \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.

```

1962 \def\@initiate@active@char#1{%
1963   \bbl@ifunset{active@char\string#1}%
1964   {\bbl@withactive
1965     {\expandafter\@initiate@active@char\expandafter}#1\string#1#1}%
1966   {}}

```

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).

```

1967 \def\@initiate@active@char#1#2#3{%
1968   \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
1969   \ifx#1\@undefined
1970     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
1971   \else
1972     \bbl@csarg\let{oridef@#2}#1%
1973     \bbl@csarg\edef{oridef@#2}{%
1974       \let\noexpand#1%
1975       \expandafter\noexpand\csname bbl@oridef@#2\endcsname}%
1976   \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*").

```

1977   \ifx#1#3\relax
1978     \expandafter\let\csname normal@char#2\endcsname#3%
1979   \else
1980     \bbl@info{Making #2 an active character}%
1981     \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
1982     \namedef{normal@char#2}{%
1983       \textormath{#3}{\csname bbl@oridef@#2\endcsname}}%
1984   \else
1985     \namedef{normal@char#2}{#3}%

```


1986 \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).

```

1987     \bbl@restoreactive{#2}%
1988     \AtBeginDocument{%
1989         \catcode`#2\active
1990         \if@filesw
1991             \immediate\write\@mainaux{\catcode`\string#2\active}%
1992         \fi}%
1993     \expandafter\bbl@add@special\csname#2\endcsname
1994     \catcode`#2\active
1995     \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⟩).

```

1996     \let\bbl@tempa\@firstoftwo
1997     \if\string^#2%
1998         \def\bbl@tempa{\noexpand\textormath}%
1999     \else
2000         \ifx\bbl@mathnormal\@undefined\else
2001             \let\bbl@tempa\bbl@mathnormal
2002         \fi
2003     \fi
2004     \expandafter\edef\csname active@char#2\endcsname{%
2005         \bbl@tempa
2006         {\noexpand\if@safe@actives
2007             \noexpand\expandafter
2008             \expandafter\noexpand\csname normal@char#2\endcsname
2009             \noexpand\else
2010             \noexpand\expandafter
2011             \expandafter\noexpand\csname bbl@doactive#2\endcsname
2012             \noexpand\fi}%
2013         {\expandafter\noexpand\csname normal@char#2\endcsname}}%
2014     \bbl@csarg\edef{doactive#2}{%
2015         \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!).

```

2016     \bbl@csarg\edef{active@#2}{%
2017         \noexpand\active@prefix\noexpand#1%
2018         \expandafter\noexpand\csname active@char#2\endcsname}%
2019     \bbl@csarg\edef{normal@#2}{%
2020         \noexpand\active@prefix\noexpand#1%
2021         \expandafter\noexpand\csname normal@char#2\endcsname}%
2022     \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.

```
2023 \bbl@active@def#2\user@group{user@active}{language@active}%
2024 \bbl@active@def#2\language@group{language@active}{system@active}%
2025 \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.

```
2026 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
2027 {\expandafter\noexpand\csname normal@char#2\endcsname}%
2028 \expandafter\edef\csname\user@group @sh@#2@string\protect\endcsname
2029 {\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.

```
2030 \if\string'#2%
2031 \let\prim@s\bbl@prim@s
2032 \let\active@math@prime#1%
2033 \fi
2034 \bbl@usehooks{initiateactive}{\{#1\}{#2\}{#3\}}
```

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

```
2035 <<*More package options>> ≡
2036 \DeclareOption{math=active}{}
2037 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
2038 <</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*.

```
2039 \@ifpackagewith{babel}{KeepShorthandsActive}%
2040 {\let\bbl@restoreactive\@gobble}%
2041 {\def\bbl@restoreactive#1{%
2042 \bbl@exp{%
2043 \\\AfterBabelLanguage\\CurrentOption
2044 {\catcode`#1=\the\catcode`#1\relax}%
2045 \\\AtEndOfPackage
2046 {\catcode`#1=\the\catcode`#1\relax}}}%
2047 \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.

```
2048 \def\bbl@sh@select#1#2{%
2049 \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
2050 \bbl@afterelse\bbl@scndcs
2051 \else
```

```

2052     \bbl@afterfi\csname#1@sh@#2@sel\endcsname
2053     \fi}

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

2054 \begingroup
2055 \bbl@ifunset{ifincsname}% TODO. Ugly. Correct?
2056 {\gdef\active@prefix#1{%
2057     \ifx\protect\@typeset@protect
2058     \else
2059         \ifx\protect\@unexpandable@protect
2060             \noexpand#1%
2061         \else
2062             \protect#1%
2063         \fi
2064     \expandafter\@gobble
2065     \fi}}
2066 {\gdef\active@prefix#1{%
2067     \ifincsname
2068         \string#1%
2069     \expandafter\@gobble
2070     \else
2071         \ifx\protect\@typeset@protect
2072         \else
2073             \ifx\protect\@unexpandable@protect
2074                 \noexpand#1%
2075             \else
2076                 \protect#1%
2077             \fi
2078         \expandafter\expandafter\expandafter\@gobble
2079         \fi
2080     \fi}}
2081 \endgroup

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

2082 \newif\if@safe@actives
2083 \@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.

2084 \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
\bbl@deactivate 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.

2085 \def\bbl@activate#1{%
2086     \bbl@withactive{\expandafter\let\expandafter}#1%
2087     \csname bbl@active@\string#1\endcsname}
2088 \def\bbl@deactivate#1{%
2089     \bbl@withactive{\expandafter\let\expandafter}#1%
2090     \csname bbl@normal@\string#1\endcsname}

```

`\bbl@firstcs` These macros are used only as a trick when declaring shorthands.

`\bbl@scndcs` 2091 `\def\bbl@firstcs#1#2{\csname#1\endcsname}`
 2092 `\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.

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

```

2093 \def\babel@texpdf#1#2#3#4{%
2094   \ifx\texorpdfstring\undefined
2095     \textormath{#1}{#2}%
2096   \else
2097     \texorpdfstring{\textormath{#1}{#3}}{#2}%
2098     % \texorpdfstring{\textormath{#1}{#3}}{\textormath{#2}{#4}}%
2099   \fi}
2100 %
2101 \def\declare@shorthand#1#2{\@decl@short{#1}#2\nil}
2102 \def\@decl@short#1#2#3\nil#4{%
2103   \def\bbl@tempa{#3}%
2104   \ifx\bbl@tempa\empty
2105     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
2106     \bbl@ifunset{#1@sh@\string#2@}{}%
2107     {\def\bbl@tempa{#4}%
2108      \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
2109      \else
2110        \bbl@info
2111          {Redefining #1 shorthand \string#2\}%
2112          in language \CurrentOption}%
2113      \fi}%
2114   \@namedef{#1@sh@\string#2@}{#4}%
2115   \else
2116     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
2117     \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
2118     {\def\bbl@tempa{#4}%
2119      \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
2120      \else
2121        \bbl@info
2122          {Redefining #1 shorthand \string#2\string#3\}%
2123          in language \CurrentOption}%
2124      \fi}%
2125     \@namedef{#1@sh@\string#2@\string#3@}{#4}%
2126   \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.

```

2127 \def\textormath{%
2128   \ifmmode

```

```

2129 \expandafter\@secondoftwo
2130 \else
2131 \expandafter\@firstoftwo
2132 \fi}

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

2133 \def\user@group{user}
2134 \def\language@group{english} % TODO. I don't like defaults
2135 \def\system@group{system}

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

2136 \def\useshorthands{%
2137 \ifstar\bbl@usesh@s{\bbl@usesh@x{}}
2138 \def\bbl@usesh@s#1{%
2139 \bbl@usesh@x
2140 {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bbl@activate{#1}}}%
2141 {#1}}
2142 \def\bbl@usesh@x#1#2{%
2143 \bbl@ifshorthand{#2}%
2144 {\def\user@group{user}%
2145 \initiate@active@char{#2}%
2146 #1%
2147 \bbl@activate{#2}}%
2148 {\bbl@error
2149 {Cannot declare a shorthand turned off (\string#2)}
2150 {Sorry, but you cannot use shorthands which have been\\%
2151 turned off in the package options}}}

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

2152 \def\user@language@group{user@\language@group}
2153 \def\bbl@set@user@generic#1#2{%
2154 \bbl@ifunset{user@generic@active#1}%
2155 {\bbl@active@def#1\user@language@group{user@active}{user@generic@active}%
2156 \bbl@active@def#1\user@group{user@generic@active}{language@active}%
2157 \expandafter\edef\csname#2@sh@#1@@\endcsname{%
2158 \expandafter\noexpand\csname normal@char#1\endcsname}%
2159 \expandafter\edef\csname#2@sh@#1@\string\protect@\endcsname{%
2160 \expandafter\noexpand\csname user@active#1\endcsname}}%
2161 \@empty}
2162 \newcommand\defineshorthand[3][user]{%
2163 \edef\bbl@tempa{\zap@space#1 \@empty}%
2164 \bbl@for\bbl@tempb\bbl@tempa{%
2165 \if*\expandafter\@car\bbl@tempb\@nil
2166 \edef\bbl@tempb{user@\expandafter\@gobble\bbl@tempb}%
2167 \@expandtwoargs
2168 \bbl@set@user@generic{\expandafter\string\@car#2\@nil}\bbl@tempb
2169 \fi
2170 \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 to fix it in the same way languages names are fixed. [TODO].

```
2171 \def\languageshorthands#1{\def\language@group{#1}}
```

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

```
2172 \def\aliasshorthand#1#2{%
2173   \bbl@ifshorthand{#2}%
2174   {\expandafter\ifx\csname active@char\string#2\endcsname\relax
2175     \ifx\document\@notprerr
2176       \@notshorthand{#2}%
2177     \else
2178       \initiate@active@char{#2}%
2179       \expandafter\let\csname active@char\string#2\expandafter\endcsname
2180         \csname active@char\string#1\endcsname
2181       \expandafter\let\csname normal@char\string#2\expandafter\endcsname
2182         \csname normal@char\string#1\endcsname
2183       \bbl@activate{#2}%
2184     \fi
2185   \fi}%
2186   {\bbl@error
2187     {Cannot declare a shorthand turned off (\string#2)}
2188     {Sorry, but you cannot use shorthands which have been\\%
2189       turned off in the package options}}}
```

`\@notshorthand`

```
2190 \def\@notshorthand#1{%
2191   \bbl@error{%
2192     The character '\string #1' should be made a shorthand character;\\%
2193     add the command \string\usesshorthands\string{#1\string} to
2194     the preamble.\\%
2195     I will ignore your instruction}%
2196   {You may proceed, but expect unexpected results}}}
```

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

```
2197 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
2198 \DeclareRobustCommand*\shorthandoff{%
2199   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
2200 \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.

```
2201 \def\bbl@switch@sh#1#2{%
2202   \ifx#2\@nnil\else
2203     \bbl@ifunset{\bbl@active@\string#2}%
```

```

2204      {\bbl@error
2205        {I cannot switch '\string#2' on or off--not a shorthand}%
2206        {This character is not a shorthand. Maybe you made\\
2207          a typing mistake? I will ignore your instruction}}}%
2208      {\ifcase#1%
2209        \catcode`#212\relax
2210      \or
2211        \catcode`#2\active
2212      \or
2213        \csname bbl@oricat@\string#2\endcsname
2214        \csname bbl@oridef@\string#2\endcsname
2215      \fi}%
2216      \bbl@afterfi\bbl@switch@sh#1%
2217    \fi}

```

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

```

2218 \def\babelshorthand{\active@prefix\babelshorthand\bbl@putsh}
2219 \def\bbl@putsh#1{%
2220   \bbl@ifunset{\bbl@active@\string#1}%
2221   {\bbl@putsh@i#1\@empty\@nnil}%
2222   {\csname bbl@active@\string#1\endcsname}}
2223 \def\bbl@putsh@i#1#2\@nnil{%
2224   \csname\language@group @sh@\string#1@%
2225     \ifx\@empty#2\else\string#2\fi\endcsname}
2226 \ifx\bbl@opt@shorthands\@nnil\else
2227   \let\bbl@s@initiate@active@char\initiate@active@char
2228   \def\initiate@active@char#1{%
2229     \bbl@ifshorthand{#1}{\bbl@s@initiate@active@char{#1}}{}}
2230   \let\bbl@s@switch@sh\bbl@switch@sh
2231   \def\bbl@switch@sh#1#2{%
2232     \ifx#2\@nnil\else
2233       \bbl@afterfi
2234       \bbl@ifshorthand{#2}{\bbl@s@switch@sh#1{#2}}{\bbl@switch@sh#1}%
2235     \fi}
2236   \let\bbl@s@activate\bbl@activate
2237   \def\bbl@activate#1{%
2238     \bbl@ifshorthand{#1}{\bbl@s@activate{#1}}{}}
2239   \let\bbl@s@deactivate\bbl@deactivate
2240   \def\bbl@deactivate#1{%
2241     \bbl@ifshorthand{#1}{\bbl@s@deactivate{#1}}{}}
2242 \fi

```

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

```

2243 \newcommand\ifbabelshorthand[3]{\bbl@ifunset{\bbl@active@\string#1}{#3}{#2}}

```

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

```

2244 \def\bbl@prim@s{%
2245   \prime\futurelet\@let@token\bbl@pr@m@s}
2246 \def\bbl@if@primes#1#2{%
2247   \ifx#1\@let@token
2248     \expandafter\@firstoftwo
2249   \else\ifx#2\@let@token
2250     \bbl@afterelse\expandafter\@firstoftwo

```

```

2251 \else
2252 \bbl@afterfi\expandafter\@secondoftwo
2253 \fi\fi}
2254 \begingroup
2255 \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
2256 \catcode`\'=12 \catcode`\\"=\active \lccode`\\"=\'
2257 \lowercase{%
2258 \gdef\bbl@pr@m@s{%
2259 \bbl@if@primes"%
2260 \pr@@s
2261 {\bbl@if@primes*\pr@@t\egroup}}
2262 \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).

```

2263 \initiate@active@char{~}
2264 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
2265 \bbl@activate{~}

```

\OT1dpos 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.

```

2266 \expandafter\def\csname OT1dpos\endcsname{127}
2267 \expandafter\def\csname T1dpos\endcsname{4}

```

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

```

2268 \ifx\f@encoding\@undefined
2269 \def\f@encoding{OT1}
2270 \fi

```

9.6 Language attributes

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

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

```

2271 \bbl@trace{Language attributes}
2272 \newcommand\languageattribute[2]{%
2273 \def\bbl@tempc{#1}%
2274 \bbl@fixname\bbl@tempc
2275 \bbl@iflanguage\bbl@tempc{%
2276 \bbl@vforeach{#2}{%

```

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

```

2277 \ifx\bbl@known@attribs\@undefined
2278 \in@false
2279 \else
2280 \bbl@xin@{\bbl@tempc-##1,}{,\bbl@known@attribs,}%

```



```

2281     \fi
2282     \ifin@
2283         \bbl@warning{%
2284             You have more than once selected the attribute '##1'\%
2285             for language #1. Reported}%
2286     \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 T_EX-code.

```

2287     \bbl@exp{%
2288         \\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
2289     \edef\bbl@tempa{\bbl@tempc-##1}%
2290     \expandafter\bbl@ifknown@ttrib\expandafter{\bbl@tempa}\bbl@attributes%
2291     {\csname\bbl@tempc @attr##1\endcsname}%
2292     {\@attrerr{\bbl@tempc}{##1}}%
2293     \fi}}
2294 \@onlypreamble\languageattribute

```

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

```

2295 \newcommand*{\@attrerr}[2]{%
2296     \bbl@error
2297     {The attribute #2 is unknown for language #1.}%
2298     {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}`.

```

2299 \def\bbl@declare@ttribute#1#2#3{%
2300     \bbl@xin@{,#2,}{,\BabelModifiers,}%
2301     \ifin@
2302         \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
2303     \fi
2304     \bbl@add@list\bbl@attributes{#1-#2}%
2305     \expandafter\def\csname#1@attr#2\endcsname{#3}}

```

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

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

First we need to find out if any attributes were set; if not we're done. Then we need to check the list of known attributes. 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`'.

```

2306 \def\bbl@ifattributeset#1#2#3#4{%
2307     \ifx\bbl@known@attribs\undefined
2308         \in@false
2309     \else
2310         \bbl@xin@{,#1-#2,}{,\bbl@known@attribs,}%
2311     \fi
2312     \ifin@
2313         \bbl@afterelse#3%
2314     \else

```

```

2315     \bbl@afterfi#4%
2316     \fi
2317 }

\bbl@ifknown@ttrib An internal macro to check whether a given language/attribute is known. The macro takes
                    4 arguments, the language/attribute, the attribute list, the  $\TeX$ -code to be executed when
                    the attribute is known and the  $\TeX$ -code to be executed otherwise.
                    We first assume the attribute is unknown. Then we loop over the list of known attributes,
                    trying to find a match. When a match is found the definition of \bbl@tempa is changed.
                    Finally we execute \bbl@tempa.

2318 \def\bbl@ifknown@ttrib#1#2{%
2319     \let\bbl@tempa\@secondoftwo
2320     \bbl@loopx\bbl@tempb{#2}{%
2321         \expandafter\in\expandafter{\expandafter,\bbl@tempb,}{, #1,}%
2322         \ifin@
2323             \let\bbl@tempa\@firstoftwo
2324         \else
2325             \fi}%
2326     \bbl@tempa
2327 }

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

2328 \def\bbl@clear@ttribs{%
2329     \ifx\bbl@attributes\undefined\else
2330         \bbl@loopx\bbl@tempa{\bbl@attributes}{%
2331             \expandafter\bbl@clear@ttrib\bbl@tempa.
2332             }%
2333         \let\bbl@attributes\undefined
2334     \fi}
2335 \def\bbl@clear@ttrib#1-#2.{%
2336     \expandafter\let\csname#1@attr#2\endcsname\undefined}
2337 \AtBeginDocument{\bbl@clear@ttribs}

```

9.7 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
2338 \bbl@trace{Macros for saving definitions}
2339 \def\babel@beginsave{\babel@savecnt\z@}

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

2340 \newcount\babel@savecnt
2341 \babel@beginsave

\babel@save The macro \babel@save<cname> saves the current meaning of the control sequence
\babel@savevariable <cname> to \originalTeX31. To do this, we let the current meaning to a temporary control
                    sequence, the restore commands are appended to \originalTeX and the counter is

```

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

incremented. The macro `\babel@savevariable<variable>` saves the value of the variable. `<variable>` can be anything allowed after the `\the` primitive.

```

2342 \def\babel@save#1{%
2343   \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
2344   \toks@\expandafter{\originalTeX\let#1=}%
2345   \bbl@exp{%
2346     \def\\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
2347   \advance\babel@savecnt@ne}
2348 \def\babel@savevariable#1{%
2349   \toks@\expandafter{\originalTeX #1=}%
2350   \bbl@exp{\def\\originalTeX{\the\toks@the#1\relax}}}
```

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

```

2351 \def\bbl@frenchspacing{%
2352   \ifnum\the\sfcodes\.=\@m
2353     \let\bbl@nonfrenchspacing\relax
2354   \else
2355     \frenchspacing
2356     \let\bbl@nonfrenchspacing\nonfrenchspacing
2357   \fi}
2358 \let\bbl@nonfrenchspacing\nonfrenchspacing
2359 %
2360 \let\bbl@elt\relax
2361 \edef\bbl@fs@chars{%
2362   \bbl@elt{\string.}\@m{3000}\bbl@elt{\string?}\@m{3000}%
2363   \bbl@elt{\string!}\@m{3000}\bbl@elt{\string:}\@m{2000}%
2364   \bbl@elt{\string;}\@m{1500}\bbl@elt{\string,}\@m{1250}}
```

9.8 Short tags

`\babeltags` This macro is straightforward. After zapping spaces, we loop over the list and define the
 macros `\text<tag>` and `\<tag>`. Definitions are first expanded so that they don't contain
`\csname` but the actual macro.

```

2365 \bbl@trace{Short tags}
2366 \def\babeltags#1{%
2367   \edef\bbl@tempa{\zap@space#1 \@empty}%
2368   \def\bbl@tempb##1=##2\@{#}%
2369   \edef\bbl@tempc{%
2370     \noexpand\newcommand
2371     \expandafter\noexpand\csname ##1\endcsname{%
2372       \noexpand\protect
2373       \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
2374     \noexpand\newcommand
2375     \expandafter\noexpand\csname text##1\endcsname{%
2376       \noexpand\foreignlanguage{##2}}
2377     \bbl@tempc}%
2378   \bbl@for\bbl@tempa\bbl@tempa{%
2379     \expandafter\bbl@tempb\bbl@tempa\@{#}}
```

9.9 Hyphens

`\babelhyphenation` This macro saves hyphenation exceptions. Two macros are used to store them:
`\bbl@hyphenation@` for the global ones and `\bbl@hyphenation<lang>` for language ones.

See `\bbl@patterns` above for further details. We make sure there is a space between words when multiple commands are used.

```

2380 \bbl@trace{Hyphens}
2381 \@onlypreamble\babelhyphenation
2382 \AtEndOfPackage{%
2383   \newcommand\babelhyphenation[2][\@empty]{%
2384     \ifx\bbl@hyphenation@relax
2385       \let\bbl@hyphenation@\@empty
2386     \fi
2387     \ifx\bbl@hyphlist\@empty\else
2388       \bbl@warning{%
2389         You must not intermingle \string\selectlanguage\space and\%
2390         \string\babelhyphenation\space or some exceptions will not\%
2391         be taken into account. Reported}%
2392     \fi
2393     \ifx\@empty#1%
2394       \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
2395     \else
2396       \bbl@vforeach{#1}{%
2397         \def\bbl@tempa{##1}%
2398         \bbl@fixname\bbl@tempa
2399         \bbl@iflanguage\bbl@tempa{%
2400           \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
2401             \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
2402             \@empty
2403             {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
2404             #2}}}%
2405     \fi}}

```

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

```

2406 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\zskip\fi}
2407 \def\bbl@t@one{T1}
2408 \def\allowhyphens{\ifx\cf@encoding\bbl@t@one\else\bbl@allowhyphens\fi}

```

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

```

2409 \newcommand\babelnullhyphen{\char\hyphenchar\font}
2410 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
2411 \def\bbl@hyphen{%
2412   \@ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
2413 \def\bbl@hyphen@i#1#2{%
2414   \bbl@ifunset{bbl@hy@#1#2\@empty}%
2415   {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
2416   {\csname bbl@hy@#1#2\@empty\endcsname}}

```

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

There should not be a discretionary after a hyphen at the beginning of a word, so it is prevented if preceded by a skip. Unfortunately, this does handle cases like “(-suffix)”. `\nobreak` is always preceded by `\leavevmode`, in case the shorthand starts a paragraph.

```

2417 \def\bbl@usehyphen#1{%

```

³² \TeX begins and ends a word for hyphenation at a glue node. The penalty prevents a linebreak at this glue node.

```

2418 \leavevmode
2419 \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
2420 \nobreak\hskip\z@skip}
2421 \def\bbl@usehyphen#1{%
2422 \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

2423 \def\bbl@hyphenchar{%
2424 \ifnum\hyphenchar\font=\m@ne
2425 \babeinullhyphen
2426 \else
2427 \char\hyphenchar\font
2428 \fi}

```

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

```

2429 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2430 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2431 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
2432 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}
2433 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}
2434 \def\bbl@hy@nobreak{\mbox{\bbl@hyphenchar}}
2435 \def\bbl@hy@repeat{%
2436 \bbl@usehyphen{%
2437 \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
2438 \def\bbl@hy@repeat{%
2439 \bbl@usehyphen{%
2440 \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}
2441 \def\bbl@hy@empty{\hskip\z@skip}
2442 \def\bbl@hy@empty{\discretionary{}{}{}}

```

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

```

2443 \def\bbl@disc#1#2{\nobreak\discretionary{#2-}{#1}\bbl@allowhyphens}

```

9.10 Multiencoding strings

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

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

```

2444 \bbl@trace{Multiencoding strings}
2445 \def\bbl@tglobal#1{\global\let#1#1}
2446 \def\bbl@recatcode#1{% TODO. Used only once?
2447 \@tempcnta="7F
2448 \def\bbl@tempa{%
2449 \ifnum\@tempcnta>"FF\else
2450 \catcode\@tempcnta=#1\relax
2451 \advance\@tempcnta\@ne
2452 \expandafter\bbl@tempa
2453 \fi}%
2454 \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).

```
2455 \@ifpackagewith{babel}{nocase}%
2456   {\let\bbl@patchuclc\relax}%
2457   {\def\bbl@patchuclc{%
2458     \global\let\bbl@patchuclc\relax
2459     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}%
2460     \gdef\bbl@uclc##1{%
2461       \let\bbl@encoded\bbl@encoded@uclc
2462       \bbl@ifunset{\language @bbl@uclc}% and resumes it
2463       {##1}%
2464       {\let\bbl@tempa##1\relax % Used by LANG@bbl@uclc
2465        \csname\language @bbl@uclc\endcsname}%
2466       {\bbl@tolower\@empty}{\bbl@toupper\@empty}}}%
2467   \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
2468   \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}%

2469 <<More package options>> ≡
2470 \DeclareOption{nocase}{}
2471 <</More package options>>
```

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

```
2472 <<More package options>> ≡
2473 \let\bbl@opt@strings\@nnil % accept strings=value
2474 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
2475 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
2476 \def\BabelStringsDefault{generic}
2477 <</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.

```
2478 \@onlypreamble\StartBabelCommands
2479 \def\StartBabelCommands{%
2480   \begingroup
2481   \bbl@recatcode{11}%
2482   <<Macros local to BabelCommands>>
2483   \def\bbl@provstring##1##2{%
2484     \providecommand##1{##2}%
2485     \bbl@tglobal##1}%
2486   \global\let\bbl@scafter\@empty
2487   \let\StartBabelCommands\bbl@startcmds
2488   \ifx\BabelLanguages\relax
2489     \let\BabelLanguages\CurrentOption
2490   \fi
2491   \begingroup
2492   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
2493   \StartBabelCommands}
2494 \def\bbl@startcmds{%
2495   \ifx\bbl@screset\@nnil\else
```

```

2496 \bbl@usehooks{stopcommands}{}%
2497 \fi
2498 \endgroup
2499 \begingroup
2500 \@ifstar
2501 {\ifx\bbl@opt@strings\@nnil
2502 \let\bbl@opt@strings\BabelStringsDefault
2503 \fi
2504 \bbl@startcmds@i}%
2505 \bbl@startcmds@i}
2506 \def\bbl@startcmds@i#1#2{%
2507 \edef\bbl@L{\zap@space#1 \@empty}%
2508 \edef\bbl@G{\zap@space#2 \@empty}%
2509 \bbl@startcmds@ii}
2510 \let\bbl@startcommands\StartBabelCommands

```

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

Select the behavior of \SetString. There are two main cases, depending of if there is an optional argument: without it and strings=encoded, strings are defined always; otherwise, they are set only if they are still undefined (ie, fallback values). With labelled blocks and strings=encoded, define the strings, but with another value, define strings only if the current label or font encoding is the value of strings; otherwise (ie, no strings or a block whose label is not in strings=) do nothing.

We presume the current block is not loaded, and therefore set (above) a couple of default values to gobble the arguments. Then, these macros are redefined if necessary according to several parameters.

```

2511 \newcommand\bbl@startcmds@ii[1][\@empty]{%
2512 \let\SetString@gobbletwo
2513 \let\bbl@stringdef@gobbletwo
2514 \let\AfterBabelCommands@gobble
2515 \ifx\@empty#1%
2516 \def\bbl@sc@label{generic}%
2517 \def\bbl@encstring##1##2{%
2518 \ProvideTextCommandDefault##1{##2}%
2519 \bbl@tglobal##1%
2520 \expandafter\bbl@tglobal\csname\string?\string##1\endcsname}%
2521 \let\bbl@sctest\in@true
2522 \else
2523 \let\bbl@sc@charset\space % <- zapped below
2524 \let\bbl@sc@fontenc\space % <- " "
2525 \def\bbl@tempa##1=##2\@nil{%
2526 \bbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
2527 \bbl@vforeach{label=#1}{\bbl@tempa##1\@nil}%
2528 \def\bbl@tempa##1 ##2{% space -> comma
2529 ##1%
2530 \ifx\@empty##2\else\ifx,##1,\else,\fi\bbl@afterfi\bbl@tempa##2\fi}%
2531 \edef\bbl@sc@fontenc{\expandafter\bbl@tempa\bbl@sc@fontenc\@empty}%
2532 \edef\bbl@sc@label{\expandafter\zap@space\bbl@sc@label\@empty}%
2533 \edef\bbl@sc@charset{\expandafter\zap@space\bbl@sc@charset\@empty}%
2534 \def\bbl@encstring##1##2{%
2535 \bbl@foreach\bbl@sc@fontenc{%
2536 \bbl@ifunset{T@###1}%
2537 }%
2538 {\ProvideTextCommand##1{####1}{##2}%
2539 \bbl@tglobal##1%
2540 \expandafter
2541 \bbl@tglobal\csname####1\string##1\endcsname}}}%
2542 \def\bbl@sctest{%

```

```

2543 \bbl@xin@{\bbl@opt@strings,}{\bbl@sc@label,\bbl@sc@fontenc,}}%
2544 \fi
2545 \ifx\bbl@opt@strings\@nnil % ie, no strings key -> defaults
2546 \else\ifx\bbl@opt@strings\relax % ie, strings=encoded
2547 \let\AfterBabelCommands\bbl@aftercmds
2548 \let\SetString\bbl@setstring
2549 \let\bbl@stringdef\bbl@encstring
2550 \else % ie, strings=value
2551 \bbl@sctest
2552 \ifin@
2553 \let\AfterBabelCommands\bbl@aftercmds
2554 \let\SetString\bbl@setstring
2555 \let\bbl@stringdef\bbl@provstring
2556 \fi\fi\fi
2557 \bbl@scswitch
2558 \ifx\bbl@G\@empty
2559 \def\SetString##1##2{%
2560 \bbl@error{Missing group for string \string##1}%
2561 {You must assign strings to some category, typically\\%
2562 captions or extras, but you set none}}%
2563 \fi
2564 \ifx\@empty#1%
2565 \bbl@usehooks{defaultcommands}{}%
2566 \else
2567 \@expandtwoargs
2568 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
2569 \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) .

```

2570 \def\bbl@forlang#1#2{%
2571 \bbl@for#1\bbl@L{%
2572 \bbl@xin@{,#1,}{\BabelLanguages,}%
2573 \ifin@#2\relax\fi}}
2574 \def\bbl@scswitch{%
2575 \bbl@forlang\bbl@tempa{%
2576 \ifx\bbl@G\@empty\else
2577 \ifx\SetString@gobbletwo\else
2578 \edef\bbl@GL{\bbl@G\bbl@tempa}%
2579 \bbl@xin@{\bbl@GL,}{\bbl@screset,}%
2580 \ifin@\else
2581 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
2582 \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
2583 \fi
2584 \fi
2585 \fi}}
2586 \AtEndOfPackage{%
2587 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{}{#2}}}%
2588 \let\bbl@scswitch\relax}
2589 \@onlypreamble\EndBabelCommands
2590 \def\EndBabelCommands{%

```



```

2591 \bbl@usehooks{stopcommands}{}%
2592 \endgroup
2593 \endgroup
2594 \bbl@scafter}
2595 \let\bbl@endcommands\EndBabelCommands

```

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.

```

2596 \def\bbl@setstring#1#2{% eg, \prefacename{<string>}
2597 \bbl@forlang\bbl@tempa{%
2598 \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
2599 \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
2600 {\bbl@exp{%
2601 \global\bbbl@add\<\bbl@G\bbl@tempa>{\bbbl@scset\#1\<\bbl@LC>}}}%
2602 }%
2603 \def\BabelString{#2}%
2604 \bbl@usehooks{stringprocess}{}%
2605 \expandafter\bbl@stringdef
2606 \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.

```

2607 \ifx\bbl@opt@strings\relax
2608 \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
2609 \bbl@patchuclc
2610 \let\bbl@encoded\relax
2611 \def\bbl@encoded@uclc#1{%
2612 \@inmathwarn#1%
2613 \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
2614 \expandafter\ifx\csname ?\string#1\endcsname\relax
2615 \TextSymbolUnavailable#1%
2616 \else
2617 \csname ?\string#1\endcsname
2618 \fi
2619 \else
2620 \csname\cf@encoding\string#1\endcsname
2621 \fi}
2622 \else
2623 \def\bbl@scset#1#2{\def#1{#2}}
2624 \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.

```

2625 <<(*Macros local to BabelCommands)>> ≡
2626 \def\SetStringLoop##1##2{%
2627 \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
2628 \count@\z@
2629 \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
2630 \advance\count@\@ne

```

```

2631 \toks@expandafter{\bbl@tempa}%
2632 \bbl@exp{%
2633 \\\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
2634 \count@=\the\count@relax}}}%
2635 <</Macros local to BabelCommands>>

```

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

```

2636 \def\bbl@aftercmds#1{%
2637 \toks@expandafter{\bbl@scafter#1}%
2638 \xdef\bbl@scafter{\the\toks@}}

```

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

```

2639 <<*Macros local to BabelCommands>> ≡
2640 \newcommand\SetCase[3][]{%
2641 \bbl@patchucl
2642 \bbl@forlang\bbl@tempa{%
2643 \expandafter\bbl@encstring
2644 \csname\bbl@tempa @bbl@ucl\endcsname{\bbl@tempa##1}%
2645 \expandafter\bbl@encstring
2646 \csname\bbl@tempa @bbl@uc\endcsname{##2}%
2647 \expandafter\bbl@encstring
2648 \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
2649 <</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.

```

2650 <<*Macros local to BabelCommands>> ≡
2651 \newcommand\SetHyphenMap[1]{%
2652 \bbl@forlang\bbl@tempa{%
2653 \expandafter\bbl@stringdef
2654 \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
2655 <</Macros local to BabelCommands>>

```

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

```

2656 \newcommand\BabelLower[2]{% one to one.
2657 \ifnum\lccode#1=#2\else
2658 \babel@savevariable{\lccode#1}%
2659 \lccode#1=#2relax
2660 \fi}
2661 \newcommand\BabelLowerMM[4]{% many-to-many
2662 \@tempcnta=#1relax
2663 \@tempcntb=#4relax
2664 \def\bbl@tempa{%
2665 \ifnum\@tempcnta>#2\else
2666 \expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
2667 \advance\@tempcnta#3relax
2668 \advance\@tempcntb#3relax
2669 \expandafter\bbl@tempa
2670 \fi}%
2671 \bbl@tempa}
2672 \newcommand\BabelLowerM0[4]{% many-to-one
2673 \@tempcnta=#1relax
2674 \def\bbl@tempa{%
2675 \ifnum\@tempcnta>#2\else

```

```

2676 \expandafter\BabelLower{\the\@tempcnta}{#4}%
2677 \advance\@tempcnta#3
2678 \expandafter\bb1@tempa
2679 \fi}%
2680 \bb1@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

2681 <(*More package options)> ≡
2682 \DeclareOption{hyphenmap=off}{\chardef\bb1@opt@hyphenmap\z@}
2683 \DeclareOption{hyphenmap=first}{\chardef\bb1@opt@hyphenmap\@ne}
2684 \DeclareOption{hyphenmap=select}{\chardef\bb1@opt@hyphenmap\tw@}
2685 \DeclareOption{hyphenmap=other}{\chardef\bb1@opt@hyphenmap\thr@@}
2686 \DeclareOption{hyphenmap=other*}{\chardef\bb1@opt@hyphenmap4\relax}
2687 <(/More package options)>

```

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

```

2688 \AtEndOfPackage{%
2689 \ifx\bb1@opt@hyphenmap\undefined
2690 \bb1@xin@{,}{\bb1@language@opts}%
2691 \chardef\bb1@opt@hyphenmap\ifin4\else\@ne\fi
2692 \fi}

```

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

```

2693 \newcommand\setlocalecaption{% TODO. Catch typos. What about ensure?
2694 \@ifstar\bb1@setcaption@s\bb1@setcaption@x}
2695 \def\bb1@setcaption@x#1#2#3{% language caption-name string
2696 \bb1@trim@def\bb1@tempa{#2}%
2697 \bb1@xin@{.template}{\bb1@tempa}%
2698 \ifin@
2699 \bb1@ini@captions@template{#3}{#1}%
2700 \else
2701 \edef\bb1@tempd{%
2702 \expandafter\expandafter\expandafter
2703 \strip@prefix\expandafter\meaning\csname captions#1\endcsname}%
2704 \bb1@xin@
2705 {\expandafter\string\csname #2name\endcsname}%
2706 {\bb1@tempd}%
2707 \ifin@ % Renew caption
2708 \bb1@xin@{\string\bb1@scset}{\bb1@tempd}%
2709 \ifin@
2710 \bb1@exp{%
2711 \\\bb1@ifsamestring{\bb1@tempa}{\language}%
2712 {\\\bb1@scset\<#2name>\<#1#2name>}%
2713 {}}%
2714 \else % Old way converts to new way
2715 \bb1@ifunset{#1#2name}%
2716 {\bb1@exp{%
2717 \\\bb1@add\<captions#1>\def\<#2name>\<#1#2name>}}%
2718 \\\bb1@ifsamestring{\bb1@tempa}{\language}%
2719 {\def\<#2name>\<#1#2name>}}%
2720 {}}}%
2721 {}%
2722 \fi
2723 \else
2724 \bb1@xin@{\string\bb1@scset}{\bb1@tempd}% New
2725 \ifin@ % New way
2726 \bb1@exp{%

```

```

2727      \\bbl@add\<captions#1>\{\\bbl@scset\<#2name>\<#1#2name>\}%
2728      \\bbl@ifsamestring{\bbl@tempa}{\language\name}%
2729      {\bbl@scset\<#2name>\<#1#2name>\}%
2730      }%
2731      \else % Old way, but defined in the new way
2732      \bbl@exp{%
2733      \\bbl@add\<captions#1>\{def\<#2name>\<#1#2name>\}%
2734      \\bbl@ifsamestring{\bbl@tempa}{\language\name}%
2735      {\def\<#2name>\<#1#2name>\}%
2736      }%
2737      \fi%
2738      \fi
2739      \@namedef{#1#2name}{#3}%
2740      \toks@{\expandafter\bbl@captionslist}%
2741      \bbl@exp{\\in@\<#2name>}{\the\toks@}%
2742      \ifin@%
2743      \bbl@exp{\\bbl@add\\bbl@captionslist{\<#2name>}}%
2744      \bbl@global\bbl@captionslist
2745      \fi
2746      \fi}
2747 % \def\bbl@setcaption@s#1#2#3{} % TODO. Not yet implemented

```

9.11 Macros common to a number of languages

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

```

2748 \bbl@trace{Macros related to glyphs}
2749 \def\set@low@box#1{\setbox\tw\hbox{,}\setbox\z@\hbox{#1}%
2750   \dimen\z@\ht\z@ \advance\dimen\z@ -\ht\tw@%
2751   \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.

```

2752 \def\save@sf@q#1{\leavevmode
2753   \begingroup
2754   \edef\SF{\spacefactor\the\spacefactor}#1\SF
2755   \endgroup}

```

9.12 Making glyphs available

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

9.12.1 Quotation marks

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

```

2756 \ProvideTextCommand{\quotedblbase}{OT1}{%
2757   \save@sf@q{\set@low@box{\textquotedblright\}}%
2758   \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.

```

2759 \ProvideTextCommandDefault{\quotedblbase}{%
2760   \UseTextSymbol{OT1}{\quotedblbase}}

```

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

```
2761 \ProvideTextCommand{\quotesinglbase}{OT1}{%
2762   \save@sf@q{\set@low@box{\textquoteright\}%
2763     \box\z@\kern-.04em\bb1@allowhyphens}}
```

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

```
2764 \ProvideTextCommandDefault{\quotesinglbase}{%
2765   \UseTextSymbol{OT1}{\quotesinglbase}}
```

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

```
2766 \ProvideTextCommand{\guillemetleft}{OT1}{%
2767   \ifmmode
2768     \ll
2769   \else
2770     \save@sf@q{\nobreak
2771       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
2772   \fi}
2773 \ProvideTextCommand{\guillemetright}{OT1}{%
2774   \ifmmode
2775     \gg
2776   \else
2777     \save@sf@q{\nobreak
2778       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
2779   \fi}
2780 \ProvideTextCommand{\guillemotleft}{OT1}{%
2781   \ifmmode
2782     \ll
2783   \else
2784     \save@sf@q{\nobreak
2785       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bb1@allowhyphens}%
2786   \fi}
2787 \ProvideTextCommand{\guillemotright}{OT1}{%
2788   \ifmmode
2789     \gg
2790   \else
2791     \save@sf@q{\nobreak
2792       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bb1@allowhyphens}%
2793   \fi}
```

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

```
2794 \ProvideTextCommandDefault{\guillemetleft}{%
2795   \UseTextSymbol{OT1}{\guillemetleft}}
2796 \ProvideTextCommandDefault{\guillemetright}{%
2797   \UseTextSymbol{OT1}{\guillemetright}}
2798 \ProvideTextCommandDefault{\guillemotleft}{%
2799   \UseTextSymbol{OT1}{\guillemotleft}}
2800 \ProvideTextCommandDefault{\guillemotright}{%
2801   \UseTextSymbol{OT1}{\guillemotright}}
```

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

```
2802 \ProvideTextCommand{\guilsinglleft}{OT1}{%
2803   \ifmmode
2804     <%
2805   \else
2806     \save@sf@q{\nobreak
```

```

2807 \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
2808 \fi}
2809 \ProvideTextCommand{\guilsinglright}{OT1}{%
2810 \ifmmode
2811 >%
2812 \else
2813 \save@sf@q{\nobreak
2814 \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
2815 \fi}

```

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

```

2816 \ProvideTextCommandDefault{\guilsinglleft}{%
2817 \UseTextSymbol{OT1}{\guilsinglleft}}
2818 \ProvideTextCommandDefault{\guilsinglright}{%
2819 \UseTextSymbol{OT1}{\guilsinglright}}

```

9.12.2 Letters

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

```

2820 \DeclareTextCommand{\ij}{OT1}{%
2821 i\kern-0.02em\bbl@allowhyphens j}
2822 \DeclareTextCommand{\IJ}{OT1}{%
2823 I\kern-0.02em\bbl@allowhyphens J}
2824 \DeclareTextCommand{\ij}{T1}{\char188}
2825 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

2826 \ProvideTextCommandDefault{\ij}{%
2827 \UseTextSymbol{OT1}{\ij}}
2828 \ProvideTextCommandDefault{\IJ}{%
2829 \UseTextSymbol{OT1}{\IJ}}

```

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

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

```

2830 \def\crrtic@{\hrule height0.1ex width0.3em}
2831 \def\crttic@{\hrule height0.1ex width0.33em}
2832 \def\ddj@{%
2833 \setbox0\hbox{d}\dimen@=\ht0
2834 \advance\dimen@1ex
2835 \dimen@.45\dimen@
2836 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2837 \advance\dimen@ii.5ex
2838 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
2839 \def\DDJ@{%
2840 \setbox0\hbox{D}\dimen@=.55\ht0
2841 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2842 \advance\dimen@ii.15ex % correction for the dash position
2843 \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
2844 \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
2845 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
2846 %
2847 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
2848 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```
2849 \ProvideTextCommandDefault{\dj}{%
2850   \UseTextSymbol{OT1}{\dj}}
2851 \ProvideTextCommandDefault{\DJ}{%
2852   \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.

```
2853 \DeclareTextCommand{\SS}{OT1}{SS}
2854 \ProvideTextCommandDefault{\SS}{\UseTextSymbol{OT1}{\SS}}
```

9.12.3 Shorthands for quotation marks

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

\glq The ‘german’ single quotes.

```
\grq 2855 \ProvideTextCommandDefault{\glq}{%
2856   \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}
```

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

```
2857 \ProvideTextCommand{\grq}{T1}{%
2858   \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
2859 \ProvideTextCommand{\grq}{TU}{%
2860   \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
2861 \ProvideTextCommand{\grq}{OT1}{%
2862   \save@sf@q{\kern-.0125em
2863     \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
2864     \kern.07em\relax}}
2865 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}
```

\glqq The ‘german’ double quotes.

```
\grqq 2866 \ProvideTextCommandDefault{\glqq}{%
2867   \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
```

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

```
2868 \ProvideTextCommand{\grqq}{T1}{%
2869   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2870 \ProvideTextCommand{\grqq}{TU}{%
2871   \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2872 \ProvideTextCommand{\grqq}{OT1}{%
2873   \save@sf@q{\kern-.07em
2874     \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
2875     \kern.07em\relax}}
2876 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}
```

\flq The ‘french’ single guillemets.

```
\frq 2877 \ProvideTextCommandDefault{\flq}{%
2878   \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
2879 \ProvideTextCommandDefault{\frq}{%
2880   \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
```

```

\flqq The ‘french’ double guillemets.
\frqq 2881 \ProvideTextCommandDefault{\flqq}{%
2882   \textormath{\guillemetleft}{\mbox{\guillemetleft}}}
2883 \ProvideTextCommandDefault{\frqq}{%
2884   \textormath{\guillemetright}{\mbox{\guillemetright}}}

```

9.12.4 Umlauts and tremas

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

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

```

2885 \def\umlauthigh{%
2886   \def\bbl@umlauta##1{\leavevmode\bgroup%
2887     \expandafter\accent\csname\fontencoding dqpos\endcsname
2888     ##1\bbl@allowhyphens\egroup}%
2889   \let\bbl@umlaute\bbl@umlauta}
2890 \def\umlautlow{%
2891   \def\bbl@umlauta{\protect\lower@umlaut}}
2892 \def\umlautelower{%
2893   \def\bbl@umlaute{\protect\lower@umlaut}}
2894 \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.

```

2895 \expandafter\ifx\csname U@D\endcsname\relax
2896   \csname newdimen\endcsname\U@D
2897 \fi

```

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

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of `.45ex` depends on the METAFONT parameters with which the fonts were built. (Just try out, which value will look best.) If the new x-height is too low, it is not changed. Finally we call the `\accent` primitive, reset the old x-height and insert the base character in the argument.

```

2898 \def\lower@umlaut#1{%
2899   \leavevmode\bgroup
2900   \U@D 1ex%
2901   {\setbox\z@\hbox{%
2902     \expandafter\char\csname\fontencoding dqpos\endcsname}%
2903     \dimen@ -.45ex\advance\dimen@\ht\z@
2904     \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
2905   \expandafter\accent\csname\fontencoding dqpos\endcsname
2906   \fontdimen5\font\U@D #1%
2907   \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).

```

2908 \AtBeginDocument{%
2909   \DeclareTextCompositeCommand{"}{OT1}{a}{\bbl@umlauta{a}}%
2910   \DeclareTextCompositeCommand{"}{OT1}{e}{\bbl@umlaute{e}}%
2911   \DeclareTextCompositeCommand{"}{OT1}{i}{\bbl@umlaute{i}}%
2912   \DeclareTextCompositeCommand{"}{OT1}{\i}{\bbl@umlaute{i}}%
2913   \DeclareTextCompositeCommand{"}{OT1}{o}{\bbl@umlauta{o}}%
2914   \DeclareTextCompositeCommand{"}{OT1}{u}{\bbl@umlauta{u}}%
2915   \DeclareTextCompositeCommand{"}{OT1}{A}{\bbl@umlauta{A}}%
2916   \DeclareTextCompositeCommand{"}{OT1}{E}{\bbl@umlaute{E}}%
2917   \DeclareTextCompositeCommand{"}{OT1}{I}{\bbl@umlaute{I}}%
2918   \DeclareTextCompositeCommand{"}{OT1}{O}{\bbl@umlauta{O}}%
2919   \DeclareTextCompositeCommand{"}{OT1}{U}{\bbl@umlauta{U}}

```

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

```

2920 \ifx\l@english\@undefined
2921   \chardef\l@english\z@
2922 \fi
2923 % The following is used to cancel rules in ini files (see Amharic).
2924 \ifx\l@babelnohyphens\@undefined
2925   \newlanguage\l@babelnohyphens
2926 \fi

```

9.13 Layout

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

```

2927 \bbl@trace{Bidi layout}
2928 \providecommand\IfBabelLayout[3]{#3}%
2929 \newcommand\BabelPatchSection[1]{%
2930   \@ifundefined{#1}{}{%
2931     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
2932     \@namedef{#1}{%
2933       \@ifstar{\bbl@presec@s{#1}}{%
2934         {\@dblarg{\bbl@presec@x{#1}}}}}%
2935 \def\bbl@presec@x#1[#2]#3{%
2936   \bbl@exp{%
2937     \\select@language@x{\bbl@main@language}%
2938     \\bbl@cs{sspre@#1}%
2939     \\bbl@cs{ss@#1}%
2940     [\\foreignlanguage{\language}{\unexpanded{#2}}]%
2941     {\\foreignlanguage{\language}{\unexpanded{#3}}}%
2942     \\select@language@x{\language}}}%
2943 \def\bbl@presec@s#1#2{%
2944   \bbl@exp{%
2945     \\select@language@x{\bbl@main@language}%
2946     \\bbl@cs{sspre@#1}%
2947     \\bbl@cs{ss@#1}*%
2948     {\\foreignlanguage{\language}{\unexpanded{#2}}}%
2949     \\select@language@x{\language}}}%
2950 \IfBabelLayout{sectioning}%
2951   {\BabelPatchSection{part}}%
2952   \BabelPatchSection{chapter}%

```

```

2953 \BabelPatchSection{section}%
2954 \BabelPatchSection{subsection}%
2955 \BabelPatchSection{subsubsection}%
2956 \BabelPatchSection{paragraph}%
2957 \BabelPatchSection{subparagraph}%
2958 \def\babel@toc#1{%
2959     \select@language@x{\bbl@main@language}}{}
2960 \IfBabelLayout{captions}%
2961 {\BabelPatchSection{caption}}{}

```

9.14 Load engine specific macros

```

2962 \bbl@trace{Input engine specific macros}
2963 \ifcase\bbl@engine
2964 \input txtbabel.def
2965 \or
2966 \input luababel.def
2967 \or
2968 \input xebabel.def
2969 \fi

```

9.15 Creating and modifying languages

`\babelprovide` is a general purpose tool for creating and modifying languages. It creates the language infrastructure, and loads, if requested, an ini file. It may be used in conjunction to previously loaded ldf files.

```

2970 \bbl@trace{Creating languages and reading ini files}
2971 \newcommand\babelprovide[2][]{%
2972     \let\bbl@savelangname\language
2973     \edef\bbl@savelocaleid{\the\localeid}%
2974     % Set name and locale id
2975     \edef\language{#2}%
2976     % \global\@namedef{\bbl@lcname@#2}{#2}%
2977     \bbl@id@assign
2978     \let\bbl@KVP@captions\@nil
2979     \let\bbl@KVP@date\@nil
2980     \let\bbl@KVP@import\@nil
2981     \let\bbl@KVP@main\@nil
2982     \let\bbl@KVP@script\@nil
2983     \let\bbl@KVP@language\@nil
2984     \let\bbl@KVP@hyphenrules\@nil
2985     \let\bbl@KVP@mapfont\@nil
2986     \let\bbl@KVP@maparabic\@nil
2987     \let\bbl@KVP@mapdigits\@nil
2988     \let\bbl@KVP@intraspace\@nil
2989     \let\bbl@KVP@intrapenalty\@nil
2990     \let\bbl@KVP@onchar\@nil
2991     \let\bbl@KVP@alph\@nil
2992     \let\bbl@KVP@Alph\@nil
2993     \let\bbl@KVP@labels\@nil
2994     \bbl@csarg\let{KVP@labels*}\@nil
2995     \bbl@forkv{#1}{% TODO - error handling
2996         \in@{/{}}{##1}%
2997         \ifin@
2998             \bbl@renewinikey##1\@{##2}%
2999         \else
3000             \bbl@csarg\def{KVP@##1}{##2}%
3001         \fi}%

```

```

3002 % == init ==
3003 \ifx\bb1@screset\@undefined
3004 \bb1@ldfinit
3005 \fi
3006 % == import, captions ==
3007 \ifx\bb1@KVP@import\@nil\else
3008 \bb1@exp{\bb1@ifblank{\bb1@KVP@import}}%
3009 {\ifx\bb1@initload\relax
3010 \begingroup
3011 \def\BabelBeforeIni##1##2{\gdef\bb1@KVP@import{##1}\endinput}%
3012 \bb1@input@texini{##2}%
3013 \endgroup
3014 \else
3015 \xdef\bb1@KVP@import{\bb1@initload}%
3016 \fi}%
3017 {}%
3018 \fi
3019 \ifx\bb1@KVP@captions\@nil
3020 \let\bb1@KVP@captions\bb1@KVP@import
3021 \fi
3022 % Load ini
3023 \bb1@ifunset{date#2}%
3024 {\bb1@provide@new{##2}}%
3025 {\bb1@ifblank{##1}%
3026 {\bb1@error
3027 {If you want to modify `#2' you must tell how in\\%
3028 the optional argument. See the manual for the\\%
3029 available options.}%
3030 {Use this macro as documented}}%
3031 {\bb1@provide@renew{##2}}}%
3032 % Post tasks
3033 \bb1@ifunset{\bb1@extracaps#2}%
3034 {\bb1@exp{\bb1@babelensure[exclude=\\today]{##2}}%
3035 {\toks@%
3036 \expandafter\expandafter\expandafter
3037 {\csname \bb1@extracaps#2\endcsname}%
3038 \bb1@exp{\bb1@babelensure[exclude=\\today,include=\the\toks@]{##2}}%
3039 \bb1@ifunset{\bb1@ensure@%
3040 \language}%
3041 {\bb1@exp{\%
3042 \\\DeclareRobustCommand\<\bb1@ensure@%
3043 \language>[1]{%
3044 \\\foreignlanguage{\language}%
3045 {###1}}}%
3046 {}%
3047 \bb1@exp{\%
3048 \\\bb1@toglobal\<\bb1@ensure@%
3049 \language>%
3050 \\\bb1@toglobal\<\bb1@ensure@%
3051 \language\space>%
3052 % At this point all parameters are defined if 'import'. Now we
3053 % execute some code depending on them. But what about if nothing was
3054 % imported? We just load the very basic parameters.
3055 \bb1@load@basic{##2}%
3056 % == script, language ==
3057 % Override the values from ini or defines them
3058 \ifx\bb1@KVP@script\@nil\else
3059 \bb1@csarg\edef{sname#2}{\bb1@KVP@script}%
3060 \fi
3061 \ifx\bb1@KVP@language\@nil\else
3062 \bb1@csarg\edef{lname#2}{\bb1@KVP@language}%
3063 \fi
3064 % == onchar ==
3065 \ifx\bb1@KVP@onchar\@nil\else

```

```

3061 \bbl@luahyphenate
3062 \directlua{
3063   if Babel.locale_mapped == nil then
3064     Babel.locale_mapped = true
3065     Babel.linebreaking.add_before(Babel.locale_map)
3066     Babel.loc_to_scr = {}
3067     Babel.chr_to_loc = Babel.chr_to_loc or {}
3068   end}%
3069 \bbl@xin@{ ids }{ \bbl@KVP@onchar\space}%
3070 \ifin@
3071   \ifx\bbl@starthyphens\@undefined % Needed if no explicit selection
3072     \AddBabelHook{babel-onchar}{beforestart}{{\bbl@starthyphens}}%
3073   \fi
3074   \bbl@exp{\bbl@add\bbl@starthyphens
3075     {\bbl@patterns@lua{\language}}}%
3076   % TODO - error/warning if no script
3077   \directlua{
3078     if Babel.script_blocks['\bbl@cl{sbc}'] then
3079       Babel.loc_to_scr[\the\localeid] =
3080         Babel.script_blocks['\bbl@cl{sbc}']
3081       Babel.locale_props[\the\localeid].lc = \the\localeid\space
3082       Babel.locale_props[\the\localeid].lg = \the\@nameuse{l@\language}\space
3083     end
3084   }%
3085 \fi
3086 \bbl@xin@{ fonts }{ \bbl@KVP@onchar\space}%
3087 \ifin@
3088   \bbl@ifunset{bbl@lsys@\language}{\bbl@provide@lsys{\language}}{}%
3089   \bbl@ifunset{bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
3090   \directlua{
3091     if Babel.script_blocks['\bbl@cl{sbc}'] then
3092       Babel.loc_to_scr[\the\localeid] =
3093         Babel.script_blocks['\bbl@cl{sbc}']
3094     end}%
3095   \ifx\bbl@mapselect\@undefined
3096     \AtBeginDocument{%
3097       \expandafter\bbl@add\csname selectfont \endcsname{{\bbl@mapselect}}%
3098       {\selectfont}}%
3099     \def\bbl@mapselect{%
3100       \let\bbl@mapselect\relax
3101       \edef\bbl@prefontid{\fontid\font}}%
3102     \def\bbl@mapdir##1{%
3103       {\def\language{##1}%
3104        \let\bbl@ifrestoring\@firstoftwo % To avoid font warning
3105        \bbl@switchfont
3106        \directlua{
3107          Babel.locale_props[\the\csname bbl@id@##1\endcsname]%
3108            ['\bbl@prefontid'] = \fontid\font\space}}}%
3109     \fi
3110     \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
3111   \fi
3112   % TODO - catch non-valid values
3113 \fi
3114 % == mapfont ==
3115 % For bidi texts, to switch the font based on direction
3116 \ifx\bbl@KVP@mapfont\@nil\else
3117   \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}{}%
3118   {\bbl@error{Option '\bbl@KVP@mapfont' unknown for\%
3119     mapfont. Use 'direction'.%

```

```

3120             {See the manual for details.}}}%
3121 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}}%
3122 \bbl@ifunset{\bbl@wdir@\language}\bbl@provide@dirs{\language}}}%
3123 \ifx\bbl@mapselect\undefined
3124   \AtBeginDocument{%
3125     \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}}%
3126     {\selectfont}}%
3127   \def\bbl@mapselect{%
3128     \let\bbl@mapselect\relax
3129     \edef\bbl@prefontid{\fontid\font}}%
3130   \def\bbl@mapdir##1{%
3131     {\def\language{##1}%
3132       \let\bbl@ifrestoring\@firstoftwo % avoid font warning
3133       \bbl@switchfont
3134       \directlua{Babel.fontmap
3135         [\the\csname bbl@wdir@##1\endcsname]%
3136         [\bbl@prefontid]=\fontid\font}}}%
3137   \fi
3138   \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
3139 \fi
3140 % == Line breaking: intraspace, intrapenalty ==
3141 % For CJK, East Asian, Southeast Asian, if interspace in ini
3142 \ifx\bbl@KVP@intraspace\@nil\else % We can override the ini or set
3143   \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
3144 \fi
3145 \bbl@provide@intraspace
3146 % == Line breaking: hyphenate.other.locale ==
3147 \bbl@ifunset{\bbl@hyotl@\language}}%
3148   {\bbl@csarg\bbl@replace{hyotl@\language}{ }{,}}%
3149   \bbl@startcommands*\language}%
3150   \bbl@csarg\bbl@foreach{hyotl@\language}{%
3151     \ifcase\bbl@engine
3152       \ifnum##1<257
3153         \SetHyphenMap{\BabelLower{##1}{##1}}%
3154       \fi
3155       \else
3156         \SetHyphenMap{\BabelLower{##1}{##1}}%
3157       \fi}%
3158   \bbl@endcommands}%
3159 % == Line breaking: hyphenate.other.script ==
3160 \bbl@ifunset{\bbl@hyots@\language}}%
3161   {\bbl@csarg\bbl@replace{hyots@\language}{ }{,}}%
3162   \bbl@csarg\bbl@foreach{hyots@\language}{%
3163     \ifcase\bbl@engine
3164       \ifnum##1<257
3165         \global\lccode##1=##1\relax
3166       \fi
3167     \else
3168       \global\lccode##1=##1\relax
3169     \fi}}%
3170 % == Counters: maparabic ==
3171 % Native digits, if provided in ini (TeX level, xe and lua)
3172 \ifcase\bbl@engine\else
3173   \bbl@ifunset{\bbl@dgnat@\language}}%
3174     {\expandafter\ifx\csname bbl@dgnat@\language\endcsname\@empty\else
3175       \expandafter\expandafter\expandafter
3176       \bbl@setdigits\csname bbl@dgnat@\language\endcsname
3177       \ifx\bbl@KVP@maparabic\@nil\else
3178         \ifx\bbl@latinarabic\undefined

```

```

3179         \expandafter\let\expandafter\@arabic
3180         \csname bbl@counter@\language\endcsname
3181     \else % ie, if layout=counters, which redefines \@arabic
3182         \expandafter\let\expandafter\bbl@latinarabic
3183         \csname bbl@counter@\language\endcsname
3184     \fi
3185 \fi
3186 \fi}%
3187 \fi
3188 % == Counters: mapdigits ==
3189 % Native digits (lua level).
3190 \ifodd\bbl@engine
3191     \ifx\bbl@KVP@mapdigits\@nil\else
3192         \bbl@ifunset{bbl@dgnat\language}{}%
3193         {\RequirePackage{luatexbase}%
3194         \bbl@activate@preotf
3195         \directlua{
3196             Babel = Babel or {} %%% -> presets in luabel
3197             Babel.digits_mapped = true
3198             Babel.digits = Babel.digits or {}
3199             Babel.digits[\the\localeid] =
3200                 table.pack(string.utfvalue('\bbl@cl{dgnat}'))
3201             if not Babel.numbers then
3202                 function Babel.numbers(head)
3203                     local LOCALE = luatexbase.registernumber'bbl@attr@locale'
3204                     local GLYPH = node.id'glyph'
3205                     local inmath = false
3206                     for item in node.traverse(head) do
3207                         if not inmath and item.id == GLYPH then
3208                             local temp = node.get_attribute(item, LOCALE)
3209                             if Babel.digits[temp] then
3210                                 local chr = item.char
3211                                 if chr > 47 and chr < 58 then
3212                                     item.char = Babel.digits[temp][chr-47]
3213                                 end
3214                             end
3215                         elseif item.id == node.id'math' then
3216                             inmath = (item.subtype == 0)
3217                         end
3218                     end
3219                     return head
3220                 end
3221             end
3222         } }%
3223     \fi
3224 \fi
3225 % == Counters: alph, Alph ==
3226 % What if extras<lang> contains a \babel@save\@alph? It won't be
3227 % restored correctly when exiting the language, so we ignore
3228 % this change with the \bbl@alph@saved trick.
3229 \ifx\bbl@KVP@alph\@nil\else
3230     \toks@\expandafter\expandafter\expandafter{%
3231         \csname extras\language\endcsname}%
3232     \bbl@exp{%
3233         \def\<extras\language>{%
3234             \let\\bbl@alph@saved\\@alph
3235             \the\toks@
3236             \let\\@alph\\bbl@alph@saved
3237             \\babel@save\\@alph

```

```

3238 \let\\@alph<bbl@cntr@bbl@KVP@alph @\languagename>}}%
3239 \fi
3240 \ifx\bbl@KVP@Alph@nil\else
3241 \toks@<expandafter\expandafter\expandafter{%
3242 \csname extras\languagename\endcsname}%
3243 \bbl@exp{%
3244 \def<extras\languagename>{%
3245 \let\\bbl@Alph@savd\\@Alph
3246 \the\toks@
3247 \let\\@Alph\\bbl@Alph@savd
3248 \\babel@save\\@Alph
3249 \let\\@Alph<bbl@cntr@bbl@KVP@Alph @\languagename>}}%
3250 \fi
3251 % == require.babel in ini ==
3252 % To load or reload the babel-*.tex, if require.babel in ini
3253 \ifx\bbl@beforestart\relax\else % But not in doc aux or body
3254 \bbl@ifunset{bbl@rqtex@languagename}}}%
3255 {\expandafter\ifx\csname bbl@rqtex@languagename\endcsname\empty\else
3256 \let\BabelBeforeIni@gobbletwo
3257 \chardef\atcatcode=\catcode`\@
3258 \catcode`\@=11\relax
3259 \bbl@input@texini{\bbl@cs{rqtex@languagename}}}%
3260 \catcode`\@=\atcatcode
3261 \let\atcatcode\relax
3262 \fi}%
3263 \fi
3264 % == main ==
3265 \ifx\bbl@KVP@main@nil % Restore only if not 'main'
3266 \let\languagename\bbl@savelangname
3267 \chardef\localeid\bbl@savelocaleid\relax
3268 \fi}

```

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

```

3269 \def\bbl@provide@new#1{%
3270 \@namedef{date#1}}}% marks lang exists - required by \StartBabelCommands
3271 \@namedef{extras#1}}}%
3272 \@namedef{noextras#1}}}%
3273 \bbl@startcommands*{#1}{captions}%
3274 \ifx\bbl@KVP@captions@nil % and also if import, implicit
3275 \def\bbl@tempb##1{% elt for \bbl@captionslist
3276 \ifx##1\empty\else
3277 \bbl@exp{%
3278 \\SetString\\##1{%
3279 \\bbl@nocaption{\bbl@stripslash##1}{#1\bbl@stripslash##1}}}%
3280 \expandafter\bbl@tempb
3281 \fi}%
3282 \expandafter\bbl@tempb\bbl@captionslist\empty
3283 \else
3284 \ifx\bbl@initoload\relax
3285 \bbl@read@ini{\bbl@KVP@captions}0% Here letters cat = 11
3286 \else
3287 \bbl@read@ini{\bbl@initoload}0% Here all letters cat = 11
3288 \fi
3289 \bbl@after@ini
3290 \bbl@savestrings
3291 \fi
3292 \StartBabelCommands*{#1}{date}%
3293 \ifx\bbl@KVP@import@nil
3294 \bbl@exp{%

```

```

3295     \\SetString\\today{\\bbl@nocaption{today}{#1today}}}%
3296   \\else
3297     \\bbl@savetoday
3298     \\bbl@savedate
3299   \\fi
3300 \\bbl@endcommands
3301 \\bbl@load@basic{#1}%
3302 % == hyphenmins == (only if new)
3303 \\bbl@exp{%
3304   \\gdef\\<#1hyphenmins>{%
3305     {\\bbl@ifunset{bbl@lftm@#1}{2}{\\bbl@cs{lftm@#1}}}%
3306     {\\bbl@ifunset{bbl@rgtm@#1}{3}{\\bbl@cs{rgtm@#1}}}}}%
3307 % == hyphenrules ==
3308 \\bbl@provide@hyphens{#1}%
3309 % == frenchspacing == (only if new)
3310 \\bbl@ifunset{bbl@frspc@#1}{}%
3311   {\\edef\\bbl@tempa{\\bbl@cl{frspc}}}%
3312   \\edef\\bbl@tempa{\\expandafter\\car\\bbl@tempa\\nil}%
3313   \\if u\\bbl@tempa      % do nothing
3314   \\else\\if n\\bbl@tempa % non french
3315     \\expandafter\\bbl@add\\csname extras#1\\endcsname{%
3316       \\let\\bbl@elt\\bbl@fs@elt@i
3317       \\bbl@fs@chars}%
3318   \\else\\if y\\bbl@tempa % french
3319     \\expandafter\\bbl@add\\csname extras#1\\endcsname{%
3320       \\let\\bbl@elt\\bbl@fs@elt@ii
3321       \\bbl@fs@chars}%
3322   \\fi\\fi\\fi}%
3323 %
3324 \\ifx\\bbl@KVP@main\\nil\\else
3325   \\expandafter\\main@language\\expandafter{#1}%
3326 \\fi}
3327 % A couple of macros used above, to avoid hashes #####...
3328 \\def\\bbl@fs@elt@i#1#2#3{%
3329   \\ifnum\\sfcode`#1=#2\\relax
3330     \\babel@savevariable{\\sfcode`#1}%
3331     \\sfcode`#1=#3\\relax
3332   \\fi}%
3333 \\def\\bbl@fs@elt@ii#1#2#3{%
3334   \\ifnum\\sfcode`#1=#3\\relax
3335     \\babel@savevariable{\\sfcode`#1}%
3336     \\sfcode`#1=#2\\relax
3337   \\fi}%
3338 %
3339 \\def\\bbl@provide@renew#1{%
3340   \\ifx\\bbl@KVP@captions\\nil\\else
3341     \\StartBabelCommands*{#1}{captions}%
3342     \\bbl@read@ini{\\bbl@KVP@captions}0%   Here all letters cat = 11
3343     \\bbl@after@ini
3344     \\bbl@savestrings
3345     \\EndBabelCommands
3346   \\fi
3347   \\ifx\\bbl@KVP@import\\nil\\else
3348     \\StartBabelCommands*{#1}{date}%
3349     \\bbl@savetoday
3350     \\bbl@savedate
3351     \\EndBabelCommands
3352   \\fi
3353   % == hyphenrules ==

```



```

3354 \bbl@provide@hyphens{#1}%
3355 % Load the basic parameters (ids, typography, counters, and a few
3356 % more), while captions and dates are left out. But it may happen some
3357 % data has been loaded before automatically, so we first discard the
3358 % saved values.
3359 \def\bbl@linebreak@export{%
3360 \bbl@exportkey{lnbrk}{typography.linebreaking}{h}%
3361 \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
3362 \bbl@exportkey{lfthm}{typography.lefthyphenmin}{2}%
3363 \bbl@exportkey{rgthm}{typography.righthyphenmin}{3}%
3364 \bbl@exportkey{prehc}{typography.prehyphenchar}{}%
3365 \bbl@exportkey{hyotl}{typography.hyphenate.other.locale}{}%
3366 \bbl@exportkey{hyots}{typography.hyphenate.other.script}{}%
3367 \bbl@exportkey{intsp}{typography.intraspace}{}%
3368 \bbl@exportkey{chrng}{characters.ranges}{}}
3369 \def\bbl@load@basic#1{%
3370 \bbl@ifunset{bbl@inidata@\language}\language}%
3371 {\getlocaleproperty\bbl@tempa{\language}{identification/load.level}%
3372 \ifcase\bbl@tempa\else
3373 \bbl@csarg\let{lname@\language}\relax
3374 \fi}%
3375 \bbl@ifunset{bbl@lname@#1}%
3376 {\def\BabelBeforeIni##1##2{%
3377 \begingroup
3378 \let\bbl@ini@captions@aux@\gobbletwo
3379 \def\bbl@inidate ###1.###2.###3.###4\relax ###5###6}%
3380 \bbl@read@ini{##1}0%
3381 \bbl@linebreak@export
3382 \bbl@exportkey{dgnat}{numbers.digits.native}{}%
3383 \bbl@exportkey{frspc}{typography.frenchspacing}{u}% unset
3384 \ifx\bbl@initoload\relax\endinput\fi
3385 \endgroup}%
3386 \begingroup % boxed, to avoid extra spaces:
3387 \ifx\bbl@initoload\relax
3388 \bbl@input@texini{##1}%
3389 \else
3390 \setbox\z@\hbox{\BabelBeforeIni{\bbl@initoload}}}%
3391 \fi
3392 \endgroup}%
3393 {}%

```

The hyphenrules option is handled with an auxiliary macro.

```

3394 \def\bbl@provide@hyphens#1{%
3395 \let\bbl@tempa\relax
3396 \ifx\bbl@KVP@hyphenrules\@nil\else
3397 \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
3398 \bbl@foreach\bbl@KVP@hyphenrules{%
3399 \ifx\bbl@tempa\relax % if not yet found
3400 \bbl@ifsamestring{##1}{+}%
3401 {\bbl@exp{\addlanguage\<l@##1>}}}%
3402 {}%
3403 \bbl@ifunset{l@##1}%
3404 {}%
3405 {\bbl@exp{\let\bbl@tempa\<l@##1>}}}%
3406 \fi}%
3407 \fi
3408 \ifx\bbl@tempa\relax % if no opt or no language in opt found
3409 \ifx\bbl@KVP@import\@nil
3410 \ifx\bbl@initoload\relax\else

```

```

3411         \bbl@exp{%                and hyphenrules is not empty
3412         \\bbl@ifblank{\bbl@cs{hyphr@#1}}%
3413         }%
3414         {\let\\bbl@tempa<l@bbl@cl{hyphr}>}}%
3415     \fi
3416 \else % if importing
3417     \bbl@exp{%                and hyphenrules is not empty
3418     \\bbl@ifblank{\bbl@cs{hyphr@#1}}%
3419     }%
3420     {\let\\bbl@tempa<l@bbl@cl{hyphr}>}}%
3421 \fi
3422 \fi
3423 \bbl@ifunset{bbl@tempa}%      ie, relax or undefined
3424 {\bbl@ifunset{l@#1}%          no hyphenrules found - fallback
3425  {\bbl@exp{\\adddialect<l@#1>\language}}%
3426  }%
3427  {\bbl@exp{\\adddialect<l@#1>\bbl@tempa}}% found in opt list or ini
3428

```

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.

```

3429 \ifx\bbl@readstream\@undefined
3430 \csname newread\endcsname\bbl@readstream
3431 \fi
3432 \def\bbl@input@texini#1{%
3433     \bbl@bsphack
3434     \bbl@exp{%
3435         \catcode`\\%=14 \catcode`\\|=0
3436         \catcode`\\{=1 \catcode`\\|=2
3437         \lowercase{\\InputIfFileExists{babel-#1.tex}{}}%
3438         \catcode`\\%=\\the\catcode`%\relax
3439         \catcode`\\|=\\the\catcode`%\relax
3440         \catcode`\\{=\\the\catcode`%\relax
3441         \catcode`\\|=\\the\catcode`%\relax}%
3442     \bbl@esphack}
3443 \def\bbl@inipreread#1=#2\@{%
3444     \bbl@trim@def\bbl@tempa{#1}% Redundant below !!
3445     \bbl@trim\toks@{#2}%
3446     % Move trims here ??
3447     \bbl@ifunset{bbl@KVP@bbl@section/\bbl@tempa}%
3448     {\bbl@exp{%
3449         \\g@addto@macro\\bbl@inidata{%
3450         \\bbl@elt{\bbl@section}{\bbl@tempa}{\the\toks@}}}%
3451     \expandafter\bbl@inireader\bbl@tempa=#2\@}%
3452     }%
3453 \def\bbl@fetch@ini#1#2{%
3454     \bbl@exp{\def\\bbl@inidata{%
3455         \\bbl@elt{identification}{tag.ini}{#1}%
3456         \\bbl@elt{identification}{load.level}{#2}}}%
3457     \openin\bbl@readstream=babel-#1.ini
3458     \ifeof\bbl@readstream
3459         \bbl@error
3460         {There is no ini file for the requested language\\%
3461         (#1). Perhaps you misspelled it or your installation\\%
3462         is not complete.}%
3463         {Fix the name or reinstall babel.}%
3464     \else
3465         \catcode`\[=12 \catcode`\]=12 \catcode`\==12 \catcode`\&=12
3466         \catcode`\;=12 \catcode`\|=12 \catcode`\%=14 \catcode`\-=12

```

```

3467 \bbl@info{Importing
3468 \ifcase#2 \or font and identification \or basic \fi
3469 data for \language\name\%
3470 from babel-#1.ini. Reported}%
3471 \loop
3472 \if T\ifeof\bbl@readstream F\fi T\relax % Trick, because inside \loop
3473 \endlinechar\m@ne
3474 \read\bbl@readstream to \bbl@line
3475 \endlinechar\^^M
3476 \ifx\bbl@line\@empty\else
3477 \expandafter\bbl@inline\bbl@line\bbl@inline
3478 \fi
3479 \repeat
3480 \fi}
3481 \def\bbl@read@ini#1#2{%
3482 \bbl@csarg\xdef{lini@\language}{#1}%
3483 \let\bbl@section\@empty
3484 \let\bbl@savestrings\@empty
3485 \let\bbl@savetoday\@empty
3486 \let\bbl@savestate\@empty
3487 \let\bbl@inireader\bbl@iniskip
3488 \bbl@fetch@ini{#1}{#2}%
3489 \bbl@foreach\bbl@renewlist{%
3490 \bbl@ifunset{\bbl@renew##1}{\bbl@inisec[##1]\@}%}
3491 \global\let\bbl@renewlist\@empty
3492 % Ends last section. See \bbl@inisec
3493 \def\bbl@elt##1##2{\bbl@inireader##1=##2\@}%
3494 \bbl@cs{renew@\bbl@section}%
3495 \global\bbl@csarg\let{renew@\bbl@section}\relax
3496 \bbl@cs{secpost@\bbl@section}%
3497 \bbl@csarg{\global\expandafter\let}{inidata@\language}\bbl@inidata
3498 \bbl@exp{\bbl@add@list\bbl@ini@loaded{\language}}%
3499 \bbl@to\global\bbl@ini@loaded}
3500 \def\bbl@inline#1\bbl@inline{%
3501 \@ifnextchar[\bbl@inisec{\@ifnextchar;\bbl@iniskip\bbl@inipreread}#1\@}% ]

```

The special cases for comment lines and sections are handled by the two following commands. In sections, we provide the possibility to take extra actions at the end or at the start. By default, key=val pairs are ignored. The secpost “hook” is used only by ‘identification’, while secpre only by date.gregorian.licr.

```

3502 \def\bbl@iniskip#1\@{% if starts with ;
3503 \def\bbl@inisec[#1]#2\@{% if starts with opening bracket
3504 \def\bbl@elt##1##2{%
3505 \expandafter\toks\expandafter{%
3506 \expandafter{\bbl@section}{##1}{##2}}%
3507 \bbl@exp{%
3508 \g@addto@macro\bbl@inidata{\bbl@elt\the\toks}}%
3509 \bbl@inireader##1=##2\@}%
3510 \bbl@cs{renew@\bbl@section}%
3511 \global\bbl@csarg\let{renew@\bbl@section}\relax
3512 \bbl@cs{secpost@\bbl@section}%
3513 % The previous code belongs to the previous section.
3514 % -----
3515 % Now start the current one.
3516 \in@{=date.}{#1}%
3517 \ifin@
3518 \lowercase{\def\bbl@tempa{=#1}}%
3519 \bbl@replace\bbl@tempa{=date.gregorian}{}%
3520 \bbl@replace\bbl@tempa{=date.}{}%

```

```

3521 \in@{.licr=}{#1}%
3522 \ifin@
3523 \ifcase\bb1@engine
3524 \bb1@replace\bb1@tempa{.licr=}{}%
3525 \else
3526 \let\bb1@tempa\relax
3527 \fi
3528 \fi
3529 \ifx\bb1@tempa\relax\else
3530 \bb1@replace\bb1@tempa{=}{}%
3531 \bb1@exp{%
3532 \def\<bb1@inikv@#1>####1=####2\\@{%
3533 \\bb1@inidate####1...\relax{####2}{\bb1@tempa}}}%
3534 \fi
3535 \fi
3536 \def\bb1@section{#1}%
3537 \def\bb1@elt##1##2{%
3538 \@namedef{bb1@KVP@#1/##1}{}}%
3539 \bb1@cs{renew@#1}%
3540 \bb1@cs{secpre@#1}% pre-section `hook'
3541 \bb1@ifunset{bb1@inikv@#1}%
3542 {\let\bb1@inireader\bb1@iniskip}%
3543 {\bb1@exp{\let\\bb1@inireader\<bb1@inikv@#1>}}
3544 \let\bb1@renewlist@empty
3545 \def\bb1@renewinikv#1/#2\\@#3{%
3546 \bb1@ifunset{bb1@renew@#1}%
3547 {\bb1@add@list\bb1@renewlist{#1}}%
3548 {}}%
3549 \bb1@csarg\bb1@add{renew@#1}{\bb1@elt{#2}{#3}}

```

Reads a key=val line and stores the trimmed val in \bb1@kv@<section>.<key>.

```

3550 \def\bb1@inikv#1=#2\\@{% key=value
3551 \bb1@trim\def\bb1@tempa{#1}%
3552 \bb1@trim\toks@{#2}%
3553 \bb1@csarg\edef{kv@\bb1@section.\bb1@tempa}{\the\toks@}}

```

The previous assignments are local, so we need to export them. If the value is empty, we can provide a default value.

```

3554 \def\bb1@exportkey#1#2#3{%
3555 \bb1@ifunset{bb1@kv@#2}%
3556 {\bb1@csarg\gdef{#1@\language}\{#3}}%
3557 {\xexpandafter\ifx\csname bb1@kv@#2\endcsname\empty
3558 \bb1@csarg\gdef{#1@\language}\{#3}}%
3559 \else
3560 \bb1@exp{\global\let\<bb1@#1@\language>\<bb1@kv@#2>}}%
3561 \fi}}

```

Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for identification and typography. Note \bb1@secpost@identification is called always (via \bb1@inisec), while \bb1@after@ini must be called explicitly after \bb1@read@ini if necessary.

```

3562 \def\bb1@iniwarning#1{%
3563 \bb1@ifunset{bb1@kv@identification.warning#1}{}%
3564 {\bb1@warning{%
3565 From babel-\bb1@cs{lini@\language}.ini:\\
3566 \bb1@cs{kv@identification.warning#1}\\
3567 Reported }}}
3568 %
3569 \let\bb1@inikv@identification\bb1@inikv

```

```

3570 \def\bbl@secpost@identification{%
3571   \bbl@iniwarning{%
3572     \ifcase\bbl@engine
3573       \bbl@iniwarning{.pdflatex}%
3574     \or
3575       \bbl@iniwarning{.lualatex}%
3576     \or
3577       \bbl@iniwarning{.xelatex}%
3578     \fi%
3579     \bbl@exportkey{elname}{identification.name.english}{}%
3580     \bbl@exp{\bbl@exportkey{lname}{identification.name.opentype}%
3581       {\csname bbl@elname@\language\endcsname}}%
3582     \bbl@exportkey{tbcpl}{identification.tag.bcp47}{}%
3583     \bbl@exportkey{lbcpl}{identification.language.tag.bcp47}{}%
3584     \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
3585     \bbl@exportkey{esname}{identification.script.name}{}%
3586     \bbl@exp{\bbl@exportkey{sname}{identification.script.name.opentype}%
3587       {\csname bbl@esname@\language\endcsname}}%
3588     \bbl@exportkey{sbcpl}{identification.script.tag.bcp47}{}%
3589     \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}%
3590     \ifbbl@bcptoname
3591       \bbl@csarg\xdef{bcp@map@\bbl@cl{tbcpl}}{\language}%
3592     \fi}

```

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

```

3593 \let\bbl@inikv@typography\bbl@inikv
3594 \let\bbl@inikv@characters\bbl@inikv
3595 \let\bbl@inikv@numbers\bbl@inikv

```

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

```

3596 \def\bbl@inikv@counters#1=#2\@@{%
3597   \bbl@ifsamestring{#1}{digits}%
3598   {\bbl@error{The counter name 'digits' is reserved for mapping\%
3599     decimal digits}%
3600     {Use another name.}}%
3601   }%
3602   \def\bbl@tempc{#1}%
3603   \bbl@trim@def{\bbl@tempb*}{#2}%
3604   \in@{.1$}{#1$}%
3605   \ifin@
3606     \bbl@replace\bbl@tempc{.1}{}%
3607     \bbl@csarg\protected@xdef{cntr@\bbl@tempc @\language}{%
3608       \noexpand\bbl@alphanumeric{\bbl@tempc}}%
3609   \fi
3610   \in@{.F.}{#1}%
3611   \ifin@else\in@{.S.}{#1}\fi
3612   \ifin@
3613     \bbl@csarg\protected@xdef{cntr@#1@\language}{\bbl@tempb*}%
3614   \else
3615     \toks@{}% Required by \bbl@buildifcase, which returns \bbl@tempa
3616     \expandafter\bbl@buildifcase\bbl@tempb* \ \ % Space after \
3617     \bbl@csarg{\global\expandafter\let}{cntr@#1@\language}\bbl@tempa
3618   \fi}
3619 \def\bbl@after@ini{%
3620   \bbl@linebreak@export
3621   \bbl@exportkey{dgnat}{numbers.digits.native}{}%
3622   \bbl@exportkey{rqtex}{identification.require.babel}{}%

```

```

3623 \bbl@exportkey{frspc}{typography.frenchspacing}{u}% unset
3624 \bbl@tglobal\bbl@savetoday
3625 \bbl@tglobal\bbl@savestate}

```

Now captions and captions.licr, depending on the engine. And below also for dates. They rely on a few auxiliary macros. It is expected the ini file provides the complete set in Unicode and LICR, in that order.

```

3626 \ifcase\bbl@engine
3627 \bbl@csarg\def{inikv@captions.licr}#1=#2\@@{%
3628 \bbl@ini@captions@aux{#1}{#2}}
3629 \else
3630 \def\bbl@inikv@captions#1=#2\@@{%
3631 \bbl@ini@captions@aux{#1}{#2}}
3632 \fi

```

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

```

3633 \def\bbl@ini@captions@template#1#2{% string language tempa=capt-name
3634 \bbl@replace\bbl@tempa{.template}{}}%
3635 \def\bbl@toreplace{#1}{}}%
3636 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace{}}%
3637 \bbl@replace\bbl@toreplace{[ ]}{\csname}%
3638 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3639 \bbl@replace\bbl@toreplace{[ ]}{name\endcsname{}}}%
3640 \bbl@replace\bbl@toreplace{[ ]}{\endcsname{}}}%
3641 \bbl@xin@{,\bbl@tempa,}{,chapter,}% TODO. cypypaste pattern
3642 \ifin@
3643 \bbl@patchchapter
3644 \global\bbl@csarg\let{chapfmt@#2}\bbl@toreplace
3645 \fi
3646 \bbl@xin@{,\bbl@tempa,}{,appendix,}%
3647 \ifin@
3648 \bbl@patchchapter
3649 \global\bbl@csarg\let{appxfmt@#2}\bbl@toreplace
3650 \fi
3651 \bbl@xin@{,\bbl@tempa,}{,part,}%
3652 \ifin@
3653 \bbl@patchpart
3654 \global\bbl@csarg\let{partfmt@#2}\bbl@toreplace
3655 \fi
3656 \bbl@xin@{,\bbl@tempa,}{,figure,table,}%
3657 \ifin@
3658 \toks@{\expandafter{\bbl@toreplace}}%
3659 \bbl@exp{\gdef\<fnum@\bbl@tempa>{\the\toks@}}%
3660 \fi}
3661 \def\bbl@ini@captions@aux#1#2{%
3662 \bbl@trim\def\bbl@tempa{#1}%
3663 \bbl@xin@{.template}{\bbl@tempa}%
3664 \ifin@
3665 \bbl@ini@captions@template{#2}\languagename
3666 \else
3667 \bbl@ifblank{#2}%
3668 {\bbl@exp{%
3669 \toks@{\bbl@nocaption{\bbl@tempa}{\languagename\bbl@tempa name}}}%
3670 {\bbl@trim\toks@{#2}}}%
3671 \bbl@exp{%
3672 \bbl@add\bbl@savestrings{%
3673 \SetString\<\bbl@tempa name>{\the\toks@}}}%
3674 \toks@{\expandafter{\bbl@captionslist}}%
3675 \bbl@exp{\in@{\<\bbl@tempa name>}{\the\toks@}}%

```

```

3676 \ifin@\\else
3677 \bbl@exp{%
3678 \\bbl@add\<bbl@extracaps@\language\>\<bbl@tempa name\>%
3679 \\bbl@toggle\<bbl@extracaps@\language\>%
3680 \fi
3681 \fi}

```

Labels. Captions must contain just strings, no format at all, so there is new group in ini files.

```

3682 \def\bbl@list@the{%
3683 part,chapter,section,subsection,subsubsection,paragraph,%
3684 subparagraph,enumi,enumii,enumiii,enumiv,equation,figure,%
3685 table,page,footnote,mpfootnote,mpfn}
3686 \def\bbl@map@cnt#1{% #1:roman,etc, // #2:enumi,etc
3687 \bbl@ifunset{bbl@map@#1@\language}%
3688 {\@nameuse{#1}}%
3689 {\@nameuse{bbl@map@#1@\language}}}%
3690 \def\bbl@inikv@labels#1=#2@@{%
3691 \in@{.map}{#1}%
3692 \ifin@
3693 \ifx\bbl@KVP@labels\@nil\else
3694 \bbl@xin@{ map }{ \bbl@KVP@labels\space}%
3695 \ifin@
3696 \def\bbl@tempc{#1}%
3697 \bbl@replace\bbl@tempc{.map}{}%
3698 \in@{,#2,}{,arabic,roman,Roman,alph,Alph,fnsymbol,}%
3699 \bbl@exp{%
3700 \gdef\<bbl@map@\bbl@tempc @\language\>%
3701 {\ifin@<#2>\else\\localecounter{#2}\fi}}%
3702 \bbl@foreach\bbl@list@the{%
3703 \bbl@ifunset{the##1}{}%
3704 {\bbl@exp{\let\\bbl@tempd\<the##1>}%
3705 \bbl@exp{%
3706 \\bbl@sreplace\<the##1>%
3707 {\<\bbl@tempc>{##1}}{\bbl@map@cnt{\bbl@tempc}{##1}}}%
3708 \\bbl@sreplace\<the##1>%
3709 {\<\empty @\bbl@tempc>\<c##1>{\bbl@map@cnt{\bbl@tempc}{##1}}}%
3710 \expandafter\ifx\csname the##1\endcsname\bbl@tempd\else
3711 \toks@\expandafter\expandafter\expandafter{%
3712 \csname the##1\endcsname}%
3713 \expandafter\edef\csname the##1\endcsname{{\the\toks@}}%
3714 \fi}}%
3715 \fi
3716 \fi
3717 %
3718 \else
3719 %
3720 % The following code is still under study. You can test it and make
3721 % suggestions. Eg, enumerate.2 = ([enumi]).([enumii]). It's
3722 % language dependent.
3723 \in@{enumerate.}{#1}%
3724 \ifin@
3725 \def\bbl@tempa{#1}%
3726 \bbl@replace\bbl@tempa{enumerate.}{}%
3727 \def\bbl@toreplace{#2}%
3728 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace}}%
3729 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3730 \bbl@replace\bbl@toreplace{ ]}{\endcsname}}%
3731 \toks@\expandafter{\bbl@toreplace}%

```

```

3732 \bbl@exp{%
3733   \\bbl@add<extras\language>{%
3734     \\babel@save<labelenum\romannumeral\bbl@tempa>%
3735     \def<labelenum\romannumeral\bbl@tempa>{\the\toks@}}%
3736     \\bbl@tglobal<extras\language>}%
3737 \fi
3738 \fi}

```

To show correctly some captions in a few languages, we need to patch some internal macros, because the order is hardcoded. For example, in Japanese the chapter number is surrounded by two string, while in Hungarian is placed after. These replacement works in many classes, but not all. Actually, the following lines are somewhat tentative.

```

3739 \def\bbl@chapttype{chap}
3740 \ifx\@makechapterhead\@undefined
3741 \let\bbl@patchchapter\relax
3742 \else\ifx\thechapter\@undefined
3743 \let\bbl@patchchapter\relax
3744 \else\ifx\ps@headings\@undefined
3745 \let\bbl@patchchapter\relax
3746 \else
3747 \def\bbl@patchchapter{%
3748   \global\let\bbl@patchchapter\relax
3749   \bbl@add\appendix{\def\bbl@chapttype{appx}}% Not harmful, I hope
3750   \bbl@tglobal\appendix
3751   \bbl@sreplace\ps@headings
3752     {\@chapapp\thechapter}%
3753     {\bbl@chapterformat}%
3754   \bbl@tglobal\ps@headings
3755   \bbl@sreplace\chaptermark
3756     {\@chapapp\thechapter}%
3757     {\bbl@chapterformat}%
3758   \bbl@tglobal\chaptermark
3759   \bbl@sreplace\@makechapterhead
3760     {\@chapapp\space\thechapter}%
3761     {\bbl@chapterformat}%
3762   \bbl@tglobal\@makechapterhead
3763   \gdef\bbl@chapterformat{%
3764     \bbl@ifunset{bbl\bbl@chapttype fmt@language}%
3765     {\@chapapp\space\thechapter}
3766     {\@nameuse{bbl\bbl@chapttype fmt@language}}}}
3767 \fi\fi\fi
3768 \ifx\@part\@undefined
3769 \let\bbl@patchpart\relax
3770 \else
3771 \def\bbl@patchpart{%
3772   \global\let\bbl@patchpart\relax
3773   \bbl@sreplace\@part
3774     {\partname\nobreakspace\thepart}%
3775     {\bbl@partformat}%
3776   \bbl@tglobal\@part
3777   \gdef\bbl@partformat{%
3778     \bbl@ifunset{bbl@partfmt@language}%
3779     {\partname\nobreakspace\thepart}
3780     {\@nameuse{bbl@partfmt@language}}}}
3781 \fi

```

Date. TODO. Document

```

3782 % Arguments are _not_ protected.
3783 \let\bbl@calendar\@empty

```



```

3784 \DeclareRobustCommand\localedate[1][\bbl@localedate{#1}]
3785 \def\bbl@localedate#1#2#3#4{%
3786   \begingroup
3787     \ifx\@empty#1\@empty\else
3788       \let\bbl@ld@calendar\@empty
3789       \let\bbl@ld@variant\@empty
3790       \edef\bbl@tempa{\zap@space#1 \@empty}%
3791       \def\bbl@tempb##1=##2@@{\@namedef\bbl@ld##1}{##2}}%
3792       \bbl@foreach\bbl@tempa{\bbl@tempb##1\@@}%
3793       \edef\bbl@calendar{%
3794         \bbl@ld@calendar
3795         \ifx\bbl@ld@variant\@empty\else
3796           .\bbl@ld@variant
3797         \fi}%
3798       \bbl@replace\bbl@calendar{gregorian}{}%
3799     \fi
3800     \bbl@cased
3801     {\@nameuse\bbl@date@\language @\bbl@calendar}{#2}{#3}{#4}}%
3802   \endgroup}
3803 % eg: 1=months, 2=wide, 3=1, 4=dummy, 5=value, 6=calendar
3804 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
3805   \bbl@trim@def\bbl@tempa{#1.#2}%
3806   \bbl@ifsamestring{\bbl@tempa}{months.wide}%      to savedate
3807   {\bbl@trim@def\bbl@tempa{#3}%
3808     \bbl@trim\toks@{#5}%
3809     \@temptokena\expandafter{\bbl@savestate}%
3810     \bbl@exp{% Reverse order - in ini last wins
3811       \def\\bbl@savestate{%
3812         \\SetString\<month\romannumeral\bbl@tempa#6name>{\the\toks@}%
3813         \the\@temptokena}}%
3814     {\bbl@ifsamestring{\bbl@tempa}{date.long}%      defined now
3815       {\lowercase{\def\bbl@tempb{#6}}%
3816         \bbl@trim@def\bbl@toreplace{#5}%
3817         \bbl@TG@date
3818         \bbl@ifunset\bbl@date@\language @}%
3819         {\global\bbl@csarg\let{date@\language @}\bbl@toreplace
3820           % TODO. Move to a better place.
3821           \bbl@exp{%
3822             \gdef\<\language date>{\\protect\<\language date >}%
3823             \gdef\<\language date >####1####2####3{%
3824               \\bbl@usedategrouptue
3825               \<\bbl@ensure@\language >{%
3826                 \\localedate{####1}{####2}{####3}}}%
3827             \\bbl@add\\bbl@savetoday{%
3828               \\SetString\\today{%
3829                 \<\language date>%
3830                 {\the\year}{\the\month}{\the\day}}}}}%
3831       }%
3832     \ifx\bbl@tempb\@empty\else
3833       \global\bbl@csarg\let{date@\language @}\bbl@tempb\bbl@toreplace
3834     \fi}%
3835   {}}}

```

Dates will require some macros for the basic formatting. They may be redefined by language, so “semi-public” names (camel case) are used. Oddly enough, the CLDR places particles like “de” inconsistently in either in the date or in the month name.

```

3836 \let\bbl@calendar\@empty
3837 \newcommand\BabelDateSpace{\nobreakspace}
3838 \newcommand\BabelDateDot{.\@} % TODO. \let instead of repeating

```

```

3839 \newcommand\BabelDated[1]{\number#1}
3840 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
3841 \newcommand\BabelDateM[1]{\number#1}
3842 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
3843 \newcommand\BabelDateMMMM[1]{%
3844   \csname month\romannumeral#1\bb1@calendar name\endcsname}%
3845 \newcommand\BabelDatey[1]{\number#1}%
3846 \newcommand\BabelDateyy[1]{%
3847   \ifnum#1<10 0\number#1 %
3848   \else\ifnum#1<100 \number#1 %
3849   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
3850   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
3851   \else
3852     \bb1@error
3853     {Currently two-digit years are restricted to the\
3854       range 0-9999.}%
3855     {There is little you can do. Sorry.}%
3856   \fi\fi\fi\fi}
3857 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
3858 \def\bb1@replace@finish@iii#1{%
3859   \bb1@exp{\def\#1###1###2###3{\the\toks@}}
3860 \def\bb1@TG@date{%
3861   \bb1@replace\bb1@toreplace{[ ]}{\BabelDateSpace{}}%
3862   \bb1@replace\bb1@toreplace{.}{\BabelDateDot{}}%
3863   \bb1@replace\bb1@toreplace{[d]}{\BabelDated{###3}}%
3864   \bb1@replace\bb1@toreplace{[dd]}{\BabelDatedd{###3}}%
3865   \bb1@replace\bb1@toreplace{[M]}{\BabelDateM{###2}}%
3866   \bb1@replace\bb1@toreplace{[MM]}{\BabelDateMM{###2}}%
3867   \bb1@replace\bb1@toreplace{[MMMM]}{\BabelDateMMMM{###2}}%
3868   \bb1@replace\bb1@toreplace{[y]}{\BabelDatey{###1}}%
3869   \bb1@replace\bb1@toreplace{[yy]}{\BabelDateyy{###1}}%
3870   \bb1@replace\bb1@toreplace{[yyyy]}{\BabelDateyyyy{###1}}%
3871   \bb1@replace\bb1@toreplace{[y|]}{\bb1@datecctr[###1|]}%
3872   \bb1@replace\bb1@toreplace{[m|]}{\bb1@datecctr[###2|]}%
3873   \bb1@replace\bb1@toreplace{[d|]}{\bb1@datecctr[###3|]}%
3874 % Note after \bb1@replace \toks@ contains the resulting string.
3875 % TODO - Using this implicit behavior doesn't seem a good idea.
3876   \bb1@replace@finish@iii\bb1@toreplace}
3877 \def\bb1@datecctr{\expandafter\bb1@xdatecctr\expandafter}
3878 \def\bb1@xdatecctr[#1|#2]{\localenumeral{#2}{#1}}

```

Language and Script values to be used when defining a font or setting the direction are set with the following macros.

```

3879 \def\bb1@provide@lsys#1{%
3880   \bb1@ifunset{\bb1@lname@#1}%
3881     {\bb1@ini@basic{#1}}%
3882     {}%
3883   \bb1@csarg\let{lsys@#1}\@empty
3884   \bb1@ifunset{\bb1@sname@#1}{\bb1@csarg\gdef{sname@#1}{Default}}{}%
3885   \bb1@ifunset{\bb1@sotf@#1}{\bb1@csarg\gdef{sotf@#1}{DFLT}}{}%
3886   \bb1@csarg\bb1@add@list{lsys@#1}{Script=\bb1@cs{sname@#1}}%
3887   \bb1@ifunset{\bb1@lname@#1}{%
3888     {\bb1@csarg\bb1@add@list{lsys@#1}{Language=\bb1@cs{lname@#1}}}%
3889   \ifcase\bb1@engine\or\or
3890     \bb1@ifunset{\bb1@prehc@#1}{%
3891       {\bb1@exp{\bb1@ifblank{\bb1@cs{prehc@#1}}}%
3892       {}%
3893       {\ifx\bb1@xenoxyph\@undefined
3894         \let\bb1@xenoxyph\bb1@xenoxyph@d

```

```

3895 \ifx\AtBeginDocument\@notprerr
3896 \expandafter\@secondoftwo % to execute right now
3897 \fi
3898 \AtBeginDocument{%
3899 \expandafter\bb1@add
3900 \csname selectfont \endcsname{\bb1@xeno-hyph}%
3901 \expandafter\selectlanguage\expandafter{\language\language}%
3902 \expandafter\bb1@tglobal\csname selectfont \endcsname}%
3903 \fi}}%
3904 \fi
3905 \bb1@csarg\bb1@tglobal{lsys#1}}
3906 \def\bb1@xeno-hyph@d{%
3907 \bb1@ifset{\bb1@prehc@\language\language}%
3908 {\ifnum\hyphenchar\font=\default-hyphenchar
3909 \iffontchar\font\bb1@cl{prehc}\relax
3910 \hyphenchar\font\bb1@cl{prehc}\relax
3911 \else\iffontchar\font"200B
3912 \hyphenchar\font"200B
3913 \else
3914 \bb1@warning
3915 {Neither 0 nor ZERO WIDTH SPACE are available\\%
3916 in the current font, and therefore the hyphen\\%
3917 will be printed. Try changing the fontspec's\\%
3918 'HyphenChar' to another value, but be aware\\%
3919 this setting is not safe (see the manual)}%
3920 \hyphenchar\font\default-hyphenchar
3921 \fi\fi
3922 \fi}%
3923 {\hyphenchar\font\default-hyphenchar}}
3924 % \fi}

```

The following ini reader ignores everything but the identification section. It is called when a font is defined (ie, when the language is first selected) to know which script/language must be enabled. This means we must make sure a few characters are not active. The ini is not read directly, but with a proxy tex file named as the language (which means any code in it must be skipped, too).

```

3925 \def\bb1@ini@basic#1{%
3926 \def\BabelBeforeIni##1##2{%
3927 \begingroup
3928 \bb1@add\bb1@secpost@identification{\closein\bb1@readstream}%
3929 \bb1@read@ini{##1}1%
3930 \endinput % babel- .tex may contain only preamble's
3931 \endgroup}% boxed, to avoid extra spaces:
3932 {\bb1@input@texini{#1}}}

```

A tool to define the macros for native digits from the list provided in the ini file. Somewhat convoluted because there are 10 digits, but only 9 arguments in \TeX . Non-digits characters are kept. The first macro is the generic “localized” command.

```

3933 \def\bb1@setdigits#1#2#3#4#5{%
3934 \bb1@exp{%
3935 \def\<\language\language digits>####1{% ie, \langdigits
3936 \<\bb1@digits@\language\language>####1\\@nil}%
3937 \let\<\bb1@cntr@digits@\language\language>\<\language\language digits>%
3938 \def\<\language\language counter>####1{% ie, \langcounter
3939 \\\expandafter\<\bb1@counter@\language\language>%
3940 \\\csname c@####1\endcsname}%
3941 \def\<\bb1@counter@\language\language>####1{% ie, \bb1@counter@lang
3942 \\\expandafter\<\bb1@digits@\language\language>%

```



```

3993 \bbl@cs{cntr@#1.1@\language}\language#8%
3994 \ifnum#6#7#8>\z@ % TODO. An ad hoc rule for Greek. Ugly.
3995 \bbl@ifunset{bbl@cntr@#1.S.321@\language}\language}%
3996 {\bbl@cs{cntr@#1.S.321@\language}\language}%
3997 \fi}%
3998 {\bbl@cs{cntr@#1.F.\number#5#6#7#8@\language}\language}}
3999 \def\bbl@alphnum@invalid#1{%
4000 \bbl@error{Alphabetic numeral too large (#1)}%
4001 {Currently this is the limit.}}

```

The information in the identification section can be useful, so the following macro just exposes it with a user command.

```

4002 \newcommand\localeinfo[1]{%
4003 \bbl@ifunset{bbl@csname bbl@info@#1\endcsname @\language}%
4004 {\bbl@error{I've found no info for the current locale.\%
4005 The corresponding ini file has not been loaded\%
4006 Perhaps it doesn't exist}%
4007 {See the manual for details.}}%
4008 {\bbl@cs{csname bbl@info@#1\endcsname @\language}}%
4009 \@namedef{bbl@info@name.locale}{lcname}
4010 \@namedef{bbl@info@tag.ini}{lini}
4011 \@namedef{bbl@info@name.english}{elname}
4012 \@namedef{bbl@info@name.opentype}{lname}
4013 \@namedef{bbl@info@tag.bcp47}{tbc}
4014 \@namedef{bbl@info@language.tag.bcp47}{lbc}
4015 \@namedef{bbl@info@tag.opentype}{lotf}
4016 \@namedef{bbl@info@script.name}{esname}
4017 \@namedef{bbl@info@script.name.opentype}{sname}
4018 \@namedef{bbl@info@script.tag.bcp47}{sbcp}
4019 \@namedef{bbl@info@script.tag.opentype}{sotf}
4020 \let\bbl@ensureinfo\@gobble
4021 \newcommand\BabelEnsureInfo{%
4022 \ifx\InputIfFileExists\undefined\else
4023 \def\bbl@ensureinfo##1{%
4024 \bbl@ifunset{bbl@lname@##1}{\bbl@ini@basic{##1}}}%
4025 \fi
4026 \bbl@foreach\bbl@loaded{%
4027 \def\language{##1}%
4028 \bbl@ensureinfo{##1}}}%

```

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

```

4029 \newcommand\getlocaleproperty{%
4030 \@ifstar\bbl@getproperty@s\bbl@getproperty@x}
4031 \def\bbl@getproperty@s#1#2#3{%
4032 \let#1\relax
4033 \def\bbl@elt##1##2##3{%
4034 \bbl@ifsamestring{##1/##2}{##3}%
4035 {\providecommand#1{##3}%
4036 \def\bbl@elt####1####2####3{}}%
4037 {}}%
4038 \bbl@cs{inidata@#2}}%
4039 \def\bbl@getproperty@x#1#2#3{%
4040 \bbl@getproperty@s{#1}{#2}{#3}%
4041 \ifx#1\relax
4042 \bbl@error
4043 {Unknown key for locale '#2':\%
4044 #3\%

```

```

4045     \string#1 will be set to \relax}%
4046     {Perhaps you misspelled it.}%
4047 \fi}
4048 \let\bbl@ini@loaded\@empty
4049 \newcommand\LocaleForEach{\bbl@foreach\bbl@ini@loaded}

```

10 Adjusting the Babel bahavior

A generic high level inteface is provided to adjust some global and general settings.

```

4050 \newcommand\babeladjust[1]{% TODO. Error handling.
4051     \bbl@forkv{#1}{%
4052         \bbl@ifunset{bbl@ADJ@##1@##2}%
4053         {\bbl@cs{ADJ@##1}{##2}}%
4054         {\bbl@cs{ADJ@##1@##2}}}
4055 %
4056 \def\bbl@adjust@lua#1#2{%
4057     \ifvmode
4058         \ifnum\currentgrouplevel=\z@
4059             \directlua{ Babel.#2 }%
4060             \expandafter\expandafter\expandafter\@gobble
4061         \fi
4062     \fi
4063     {\bbl@error    % The error is gobbled if everything went ok.
4064         {Currently, #1 related features can be adjusted only\\%
4065             in the main vertical list.}%
4066         {Maybe things change in the future, but this is what it is.}}}
4067 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
4068     \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
4069 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
4070     \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
4071 \@namedef{bbl@ADJ@bidi.text@on}{%
4072     \bbl@adjust@lua{bidi}{bidi_enabled=true}}
4073 \@namedef{bbl@ADJ@bidi.text@off}{%
4074     \bbl@adjust@lua{bidi}{bidi_enabled=false}}
4075 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
4076     \bbl@adjust@lua{bidi}{digits_mapped=true}}
4077 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
4078     \bbl@adjust@lua{bidi}{digits_mapped=false}}
4079 %
4080 \@namedef{bbl@ADJ@linebreak.sea@on}{%
4081     \bbl@adjust@lua{linebreak}{sea_enabled=true}}
4082 \@namedef{bbl@ADJ@linebreak.sea@off}{%
4083     \bbl@adjust@lua{linebreak}{sea_enabled=false}}
4084 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
4085     \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
4086 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
4087     \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
4088 %
4089 \def\bbl@adjust@layout#1{%
4090     \ifvmode
4091         #1%
4092         \expandafter\@gobble
4093     \fi
4094     {\bbl@error    % The error is gobbled if everything went ok.
4095         {Currently, layout related features can be adjusted only\\%
4096             in vertical mode.}%
4097         {Maybe things change in the future, but this is what it is.}}}

```

```

4098 \@namedef{bbl@ADJ@layout.tabular@on}{%
4099   \bbl@adjust@layout{\let\@tabular\bbl@NL@tabular}}
4100 \@namedef{bbl@ADJ@layout.tabular@off}{%
4101   \bbl@adjust@layout{\let\@tabular\bbl@OL@tabular}}
4102 \@namedef{bbl@ADJ@layout.lists@on}{%
4103   \bbl@adjust@layout{\let\list\bbl@NL@list}}
4104 \@namedef{bbl@ADJ@layout.lists@off}{%
4105   \bbl@adjust@layout{\let\list\bbl@OL@list}}
4106 \@namedef{bbl@ADJ@hyphenation.extra@on}{%
4107   \bbl@activateposthyphen}
4108 %
4109 \@namedef{bbl@ADJ@autoload.bcp47@on}{%
4110   \bbl@bcpallowedtrue}
4111 \@namedef{bbl@ADJ@autoload.bcp47@off}{%
4112   \bbl@bcpallowedfalse}
4113 \@namedef{bbl@ADJ@autoload.bcp47.prefix}#1{%
4114   \def\bbl@bcp@prefix{#1}}
4115 \def\bbl@bcp@prefix{bcp47-}
4116 \@namedef{bbl@ADJ@autoload.options}#1{%
4117   \def\bbl@autoload@options{#1}}
4118 \let\bbl@autoload@bcptoptions\@empty
4119 \@namedef{bbl@ADJ@autoload.bcp47.options}#1{%
4120   \def\bbl@autoload@bcptoptions{#1}}
4121 \newif\ifbbl@bcptname
4122 \@namedef{bbl@ADJ@bcp47.toname@on}{%
4123   \bbl@bcptonametrue}
4124 \BabelEnsureInfo}
4125 \@namedef{bbl@ADJ@bcp47.toname@off}{%
4126   \bbl@bcptonamefalse}
4127 % TODO: use babel name, override
4128 %
4129 % As the final task, load the code for lua.
4130 %
4131 \ifx\directlua\@undefined\else
4132   \ifx\bbl@luapatterns\@undefined
4133     \input luababel.def
4134   \fi
4135 \fi
4136 \</core>

A proxy file for switch.def
4137 \<kernel>
4138 \let\bbl@onlyswitch\@empty
4139 \input babel.def
4140 \let\bbl@onlyswitch\@undefined
4141 \</kernel>
4142 \<patterns>

```

11 Loading hyphenation patterns

The following code is meant to be read by $\text{\texttt{iniT\TeX}}$ because it should instruct $\text{\texttt{T\TeX}}$ 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.

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

This new definition starts by adding an instruction to write a message on the terminal and in the transcript file to inform the user of the preloaded hyphenation patterns. Then everything is restored to the old situation and the format is dumped.

```

4143 <<Make sure ProvidesFile is defined>>
4144 \ProvidesFile{hyphen.cfg}[\<<date>>] [\<<version>>] Babel hyphens]
4145 \xdef\bbl@format{\jobname}
4146 \def\bbl@version{\<<version>>}
4147 \def\bbl@date{\<<date>>}
4148 \ifx\AtBeginDocument\@undefined
4149   \def\@empty{}
4150   \let\orig@dump\dump
4151   \def\dump{%
4152     \ifx\@ztryfc\@undefined
4153       \else
4154         \toks0=\expandafter{\@preamblecmds}%
4155         \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
4156         \def\@begindocumenthook{}%
4157       \fi
4158       \let\dump\orig@dump\let\orig@dump\@undefined\dump}
4159 \fi
4160 <<Define core switching macros>>

```

`\process@line` Each line in the file language.dat is processed by `\process@line` after it is read. The first thing this macro does is to check whether the line starts with =. When the first token of a line is an =, the macro `\process@synonym` is called; otherwise the macro `\process@language` will continue.

```

4161 \def\process@line#1#2 #3 #4 {%
4162   \ifx=#1%
4163     \process@synonym{#2}%
4164   \else
4165     \process@language{#1#2}{#3}{#4}%
4166   \fi
4167   \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.

```

4168 \toks@{}
4169 \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.

```

4170 \def\process@synonym#1{%
4171   \ifnum\last@language=\m@ne
4172     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
4173   \else
4174     \expandafter\chardef\csname l@#1\endcsname\last@language
4175     \wlog{\string\l@#1=\string\language\the\last@language}%
4176     \expandafter\let\csname #1hyphenmins\endcsname\expandafter\endcsname
4177     \csname\language\endcsname hyphenmins\endcsname
4178     \let\bbl@elt\relax
4179     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}}}%
4180   \fi}

```


`\process@language` The macro `\process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions. The first thing to do is call `\addlanguage` to allocate a pattern register and to make that register ‘active’. Then the pattern file is read.

For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file `language.dat` by adding for instance ‘:T1’ to the name of the language. The macro `\bbl@get@enc` extracts the font encoding from the language name and stores it in `\bbl@hyph@enc`. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

Pattern files may contain assignments to `\lefthyphenmin` and `\righthyphenmin`. 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.

```

4181 \def\process@language#1#2#3{%
4182   \expandafter\addlanguage\csname l@#1\endcsname
4183   \expandafter\language\csname l@#1\endcsname
4184   \edef\language#1{%
4185     \bbl@hook@everylanguage{#1}%
4186     % > luatex
4187     \bbl@get@enc#1::\@@@
4188     \begingroup
4189       \lefthyphenmin\m@ne
4190       \bbl@hook@loadpatterns{#2}%
4191       % > luatex
4192       \ifnum\lefthyphenmin=\m@ne
4193       \else
4194         \expandafter\xdef\csname #1hyphenmins\endcsname{%
4195           \the\lefthyphenmin\the\righthyphenmin}%
4196       \fi
4197     \endgroup
4198     \def\bbl@tempa{#3}%
4199     \ifx\bbl@tempa\@empty\else
4200       \bbl@hook@loadexceptions{#3}%
4201       % > luatex
4202     \fi
4203     \let\bbl@elt\relax
4204     \edef\bbl@languages{%
4205       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
4206     \ifnum\the\language=\z@

```

```

4207 \expandafter\ifx\csname #1hyphenmins\endcsname\relax
4208 \set@hyphenmins\tw@\thr@@\relax
4209 \else
4210 \expandafter\expandafter\expandafter\set@hyphenmins
4211 \csname #1hyphenmins\endcsname
4212 \fi
4213 \the\toks@
4214 \toks@{}%
4215 \fi}

```

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

```

4216 \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. loadkernel currently loads nothing, but define some basic macros instead.

```

4217 \def\bbl@hook@everylanguage#1{}
4218 \def\bbl@hook@loadpatterns#1{\input #1\relax}
4219 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
4220 \def\bbl@hook@loadkernel#1{%
4221 \def\addlanguage{\csname newlanguage\endcsname}%
4222 \def\adddialect##1##2{%
4223 \global\chardef##1##2\relax
4224 \wlog{\string##1 = a dialect from \string\language##2}}%
4225 \def\iflanguage#1{%
4226 \expandafter\ifx\csname l@##1\endcsname\relax
4227 \@nolanerr{##1}%
4228 \else
4229 \ifnum\csname l@##1\endcsname=\language
4230 \expandafter\expandafter\expandafter\@firstoftwo
4231 \else
4232 \expandafter\expandafter\expandafter\@secondoftwo
4233 \fi
4234 \fi}%
4235 \def\providehyphenmins##1##2{%
4236 \expandafter\ifx\csname ##1hyphenmins\endcsname\relax
4237 \@namedef{##1hyphenmins}{##2}%
4238 \fi}%
4239 \def\set@hyphenmins##1##2{%
4240 \lefthyphenmin##1\relax
4241 \righthyphenmin##2\relax}%
4242 \def\selectlanguage{%
4243 \errhelp{Selecting a language requires a package supporting it}%
4244 \errmessage{Not loaded}}%
4245 \let\foreignlanguage\selectlanguage
4246 \let\otherlanguage\selectlanguage
4247 \expandafter\let\csname otherlanguage*\endcsname\selectlanguage
4248 \def\bbl@usehooks##1##2{% TODO. Temporary!!
4249 \def\setlocale{%
4250 \errhelp{Find an armchair, sit down and wait}%
4251 \errmessage{Not yet available}}%
4252 \let\uselocale\setlocale
4253 \let\locale\setlocale
4254 \let\selectlocale\setlocale
4255 \let\localename\setlocale
4256 \let\textlocale\setlocale
4257 \let\textlanguage\setlocale

```

```

4258 \let\languagetext\setlocale}
4259 \begingroup
4260 \def\AddBabelHook#1#2{%
4261   \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
4262     \def\next{\toks1}%
4263   \else
4264     \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
4265   \fi
4266   \next}
4267 \ifx\directlua\undefined
4268   \ifx\XeTeXinputencoding\undefined\else
4269     \input xebabel.def
4270   \fi
4271 \else
4272   \input luababel.def
4273 \fi
4274 \openin1 = babel-\bbl@format.cfg
4275 \ifeof1
4276 \else
4277   \input babel-\bbl@format.cfg\relax
4278 \fi
4279 \closein1
4280 \endgroup
4281 \bbl@hook@loadkernel{switch.def}

```

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

```

4282 \openin1 = language.dat

```

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

```

4283 \def\languagename{english}%
4284 \ifeof1
4285   \message{I couldn't find the file language.dat,\space
4286           I will try the file hyphen.tex}
4287   \input hyphen.tex\relax
4288   \chardef\l@english\z@
4289 \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 .

```

4290 \last@language\m@ne

```

We now read lines from the file until the end is found. While reading from the input, it is useful to switch off recognition of the end-of-line character. This saves us stripping off spaces from the contents of the control sequence.

```

4291 \loop
4292   \endlinechar\m@ne
4293   \read1 to \bbl@line
4294   \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.

```

4295   \if T\ifeof1F\fi T\relax
4296   \ifx\bbl@line\@empty\else
4297     \edef\bbl@line{\bbl@line\space\space\space}%

```

```

4298      \expandafter\process@line\bbl@line\relax
4299      \fi
4300  \repeat

```

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

```

4301  \begingroup
4302      \def\bbl@elt#1#2#3#4{%
4303          \global\language=#2\relax
4304          \gdef\language{#1}%
4305          \def\bbl@elt##1##2##3##4{}}%
4306      \bbl@languages
4307  \endgroup
4308  \fi
4309  \closein1

```

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

```

4310  \if\the\toks@\else
4311      \errhelp{language.dat loads no language, only synonyms}
4312      \errmessage{Orphan language synonym}
4313  \fi

```

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

```

4314  \let\bbl@line\@undefined
4315  \let\process@line\@undefined
4316  \let\process@synonym\@undefined
4317  \let\process@language\@undefined
4318  \let\bbl@get@enc\@undefined
4319  \let\bbl@hyph@enc\@undefined
4320  \let\bbl@tempa\@undefined
4321  \let\bbl@hook@loadkernel\@undefined
4322  \let\bbl@hook@everylanguage\@undefined
4323  \let\bbl@hook@loadpatterns\@undefined
4324  \let\bbl@hook@loadexceptions\@undefined
4325  \patterns

```

Here the code for `iniTeX` ends.

12 Font handling with fontspec

Add the bidi handler just before `luaotfload`, which is loaded by default by LaTeX. Just in case, consider the possibility it has not been loaded. First, a couple of definitions related to bidi [misplaced].

```

4326  <<More package options>> ≡
4327  \chardef\bbl@bidimode\z@
4328  \DeclareOption{bidi=default}{\chardef\bbl@bidimode=\@ne}
4329  \DeclareOption{bidi=basic}{\chardef\bbl@bidimode=101 }
4330  \DeclareOption{bidi=basic-r}{\chardef\bbl@bidimode=102 }
4331  \DeclareOption{bidi=bidi}{\chardef\bbl@bidimode=201 }
4332  \DeclareOption{bidi=bidi-r}{\chardef\bbl@bidimode=202 }
4333  \DeclareOption{bidi=bidi-l}{\chardef\bbl@bidimode=203 }
4334  <</More package options>>

```

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

At the time of this writing, `fontspec` shows a warning about there are languages not available, which some people think refers to `babel`, even if there is nothing wrong. Here is hack to patch `fontspec` to avoid the misleading message, which is replaced by a more explanatory one.

```

4335 <<{*Font selection}>> ≡
4336 \bbl@trace{Font handling with fontspec}
4337 \ifx\ExplSyntaxOn\@undefined\else
4338   \ExplSyntaxOn
4339   \catcode`\ =10
4340   \def\bbl@loadfontspec{%
4341     \usepackage{fontspec}%
4342     \expandafter
4343     \def\csname msg~text~>~fontspec/language-not-exist\endcsname##1##2##3##4{%
4344       Font '\l_fontspec_fontname_tl' is using the\\%
4345       default features for language '##1'.\\%
4346       That's usually fine, because many languages\\%
4347       require no specific features, but if the output is\\%
4348       not as expected, consider selecting another font.}
4349     \expandafter
4350     \def\csname msg~text~>~fontspec/no-script\endcsname##1##2##3##4{%
4351       Font '\l_fontspec_fontname_tl' is using the\\%
4352       default features for script '##2'.\\%
4353       That's not always wrong, but if the output is\\%
4354       not as expected, consider selecting another font.}}
4355   \ExplSyntaxOff
4356 \fi
4357 \@onlypreamble\babelfont
4358 \newcommand\babelfont[2][1]{% 1=langs/scripts 2=fam
4359   \bbl@foreach{#1}{%
4360     \expandafter\ifx\csname date##1\endcsname\relax
4361       \IfFileExists{babel-##1.tex}%
4362       {\babelprovide{##1}}%
4363     }%
4364   \fi}%
4365 \edef\bbl@tempa{#1}%
4366 \def\bbl@tempb{#2}% Used by \bbl@bblfont
4367 \ifx\fontspec\@undefined
4368   \bbl@loadfontspec
4369 \fi
4370 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
4371 \bbl@bblfont}
4372 \newcommand\bbl@bblfont[2][1]{% 1=features 2=fontname, @font=rm|sf|tt
4373   \bbl@ifunset{\bbl@tempb family}%
4374   {\bbl@providefam{\bbl@tempb}}%
4375   {\bbl@exp{%
4376     \\bbl@sreplace\<\bbl@tempb family >%
4377     {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
4378   % For the default font, just in case:
4379   \bbl@ifunset{\bbl@lsys@\languagenome}{\bbl@provide@lsys{\languagenome}}}%
4380   \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
4381   {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save \bbl@rmdflt@
4382     \bbl@exp{%
4383       \let\<\bbl@tempb dflt@\languagenome>\<\bbl@tempb dflt@>%
4384       \\bbl@font@set\<\bbl@tempb dflt@\languagenome>%
4385       \<\bbl@tempb default>\<\bbl@tempb family>}}%

```

```

4386   {\bbl@foreach\bbl@tempa{% ie bbl@rmdflt@lang / *scrt
4387     \bbl@csarg\def{\bbl@tempb dflt@##1}{<{#1}{#2}}}}}%

```

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

```

4388 \def\bbl@providfam#1{%
4389   \bbl@exp{%
4390     \\newcommand\<#1default>{}% Just define it
4391     \\bbl@add@list\\bbl@font@fams{#1}%
4392     \\DeclareRobustCommand\<#1family>{%
4393       \\not@math@alphabet\<#1family>\relax
4394       \\fontfamily\<#1default>\\selectfont}%
4395     \\DeclareTextFontCommand{\<text#1>}{\<#1family>}}%

```

The following macro is activated when the hook babel-fontspec is enabled. But before we define a macro for a warning, which sets a flag to avoid duplicate them.

```

4396 \def\bbl@nostdfont#1{%
4397   \bbl@ifunset{\bbl@WFF@f@family}%
4398   {\bbl@csarg\gdef{WFF@f@family}}}% Flag, to avoid dupl warns
4399   \bbl@infowarn{The current font is not a babel standard family:\%
4400     #1%
4401     \fontname\font\\%
4402     There is nothing intrinsically wrong with this warning, and\\%
4403     you can ignore it altogether if you do not need these\\%
4404     families. But if they are used in the document, you should be\\%
4405     aware 'babel' will no set Script and Language for them, so\\%
4406     you may consider defining a new family with \string\babelfont.\\%
4407     See the manual for further details about \string\babelfont.\\%
4408     Reported}}
4409   {}}%
4410 \gdef\bbl@switchfont{%
4411   \bbl@ifunset{\bbl@lsys@\languagename}{\bbl@provide@lsys{\languagename}}}%
4412   \bbl@exp{% eg Arabic -> arabic
4413     \lowercase{\edef\\bbl@tempa{\bbl@cl{sname}}}}}%
4414   \bbl@foreach\bbl@font@fams{%
4415     \bbl@ifunset{\bbl@##1dflt@\languagename}% (1) language?
4416     {\bbl@ifunset{\bbl@##1dflt*@\bbl@tempa}% (2) from script?
4417       {\bbl@ifunset{\bbl@##1dflt@}% 2=F - (3) from generic?
4418         {}}% 123=F - nothing!
4419       {\bbl@exp{% 3=T - from generic
4420         \global\let\<bbl@##1dflt@\languagename>%
4421         \<bbl@##1dflt@>}}}%
4422       {\bbl@exp{% 2=T - from script
4423         \global\let\<bbl@##1dflt@\languagename>%
4424         \<bbl@##1dflt*@\bbl@tempa>}}}%
4425       {}}% 1=T - language, already defined
4426   \def\bbl@tempa{\bbl@nostdfont{}}}%
4427   \bbl@foreach\bbl@font@fams{% don't gather with prev for
4428     \bbl@ifunset{\bbl@##1dflt@\languagename}%
4429     {\bbl@cs{famrst@##1}%
4430     \global\bbl@csarg\let{famrst@##1}\relax}%
4431     {\bbl@exp{% order is relevant. TODO: but sometimes wrong!
4432       \\bbl@add\\originalTeX%
4433       \\bbl@font@rst{\bbl@cl{##1dflt}}}%
4434       \<##1default>\<##1family>{##1}}}%
4435       \\bbl@font@set\<bbl@##1dflt@\languagename>% the main part!
4436       \<##1default>\<##1family>}}}%
4437   \bbl@ifrestoring{}{\bbl@tempa}}%

```

The following is executed at the beginning of the aux file or the document to warn about fonts not defined with \babelfont.

```

4438 \ifx\family\undefined\else % if latex
4439 \ifcase\bbbl@engine % if pdftex
4440 \let\bbbl@ckeckstdfonts\relax
4441 \else
4442 \def\bbbl@ckeckstdfonts{%
4443 \begingroup
4444 \global\let\bbbl@ckeckstdfonts\relax
4445 \let\bbbl@tempa\@empty
4446 \bbbl@foreach\bbbl@font@fams{%
4447 \bbbl@ifunset{\bbbl@##1dflt@}%
4448 {\nameuse{##1family}%
4449 \bbbl@csarg\gdef{WFF@\family}}}% Flag
4450 \bbbl@exp{\bbbl@add\bbbl@tempa{* \<##1family>= \family\\%
4451 \space\space\fontname\font\\}%
4452 \bbbl@csarg\xdef{##1dflt@}{\family}%
4453 \expandafter\xdef\csname ##1default\endcsname{\family}}%
4454 {}}%
4455 \ifx\bbbl@tempa\@empty\else
4456 \bbbl@infowarn{The following font families will use the default\\%
4457 settings for all or some languages:\\%
4458 \bbbl@tempa
4459 There is nothing intrinsically wrong with it, but\\%
4460 'babel' will no set Script and Language, which could\\%
4461 be relevant in some languages. If your document uses\\%
4462 these families, consider redefining them with \string\babelfont.\\%
4463 Reported}%
4464 \fi
4465 \endgroup}
4466 \fi
4467 \fi

```

Now the macros defining the font with fontspec.

When there are repeated keys in fontspec, the last value wins. So, we just place the ini settings at the beginning, and user settings will take precedence. We must deactivate temporarily \bbbl@mapselect because \selectfont is called internally when a font is defined.

```

4468 \def\bbbl@font@set#1#2#3{% eg \bbbl@rmdflt@lang \rmdefault \rmfamily
4469 \bbbl@xin@{<>}{#1}%
4470 \ifin@
4471 \bbbl@exp{\bbbl@fontspec@set\#1\expandafter@gobbletwo#1\#3}%
4472 \fi
4473 \bbbl@exp{%
4474 \def\#2{#1}% eg, \rmdefault{\bbbl@rmdflt@lang}
4475 \bbbl@ifsamestring{#2}{\family}%
4476 {\#3%
4477 \bbbl@ifsamestring{\family@series}{\bfdefault}{\bfseries}}}%
4478 \let\bbbl@tempa\relax}%
4479 {}}}}
4480 % TODO - next should be global?, but even local does its job. I'm
4481 % still not sure -- must investigate:
4482 \def\bbbl@fontspec@set#1#2#3#4{% eg \bbbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
4483 \let\bbbl@tempa\bbbl@mapselect
4484 \let\bbbl@mapselect\relax
4485 \let\bbbl@tempfam#4% eg, '\rmfamily', to be restored below
4486 \let#4\@empty % Make sure \renewfontfamily is valid
4487 \bbbl@exp{%
4488 \let\bbbl@tempa\pfam\<\bbbl@stripslash#4\space>% eg, '\rmfamily '
4489 \<keys_if_exist:nnF>{fontspec-opentype}{Script/\bbbl@cl{sname}}%
4490 {\newfontscript{\bbbl@cl{sname}}{\bbbl@cl{sotf}}}%

```

```

4491 \<keys_if_exist:nnF>{fontspec-opentype}{Language/\bbl@cl{lname}}}%
4492 {\newfontlanguage{\bbl@cl{lname}}{\bbl@cl{lotf}}}%
4493 \renewfontfamily\#4%
4494 [\bbl@cs{lsys@{language},#2}]{#3}% ie \bbl@exp{.}{#3}
4495 \begingroup
4496 #4%
4497 \xdef#1{\f@family}% eg, \bbl@rmdflt@lang{FreeSerif(0)}
4498 \endgroup
4499 \let#4\bbl@temp@fam
4500 \bbl@exp{\let\<\bbl@stripslash#4\space>}\bbl@temp@pfam
4501 \let\bbl@mapselect\bbl@tempe}%

```

font@rst and famrst are only used when there is no global settings, to save and restore de previous families. Not really necessary, but done for optimization.

```

4502 \def\bbl@font@rst#1#2#3#4{%
4503 \bbl@csarg\def{famrst@#4}{\bbl@font@set{#1}#2#3}}

```

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

```

4504 \def\bbl@font@fams{rm,sf,tt}

```

The old tentative way. Short and preverved for compatibility, but deprecated. Note there is no direct alternative for \babelFSfeatures. The reason in explained in the user guide, but essentially – that was not the way to go :-).

```

4505 \newcommand\babelFSstore[2][{}]{%
4506 \bbl@ifblank{#1}%
4507 {\bbl@csarg\def{sname@#2}{Latin}}%
4508 {\bbl@csarg\def{sname@#2}{#1}}%
4509 \bbl@provide@dirs{#2}%
4510 \bbl@csarg\ifnum{wdir@#2}>\z@
4511 \let\bbl@beforeforeign\leavevmode
4512 \EnableBabelHook{babel-bidi}%
4513 \fi
4514 \bbl@foreach{#2}{%
4515 \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
4516 \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
4517 \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
4518 \def\bbl@FSstore#1#2#3#4{%
4519 \bbl@csarg\edef{#2default#1}{#3}%
4520 \expandafter\addto\csname extras#1\endcsname{%
4521 \let#4#3%
4522 \ifx#3\f@family
4523 \edef#3{\csname bbl@#2default#1\endcsname}%
4524 \fontfamily{#3}\selectfont
4525 \else
4526 \edef#3{\csname bbl@#2default#1\endcsname}%
4527 \fi}%
4528 \expandafter\addto\csname noextras#1\endcsname{%
4529 \ifx#3\f@family
4530 \fontfamily{#4}\selectfont
4531 \fi
4532 \let#3#4}}
4533 \let\bbl@langfeatures\empty
4534 \def\babelFSfeatures{% make sure \fontspec is redefined once
4535 \let\bbl@ori@fontspec\fontspec
4536 \renewcommand\fontspec[1][{}]{%
4537 \bbl@ori@fontspec[\bbl@langfeatures##1]}
4538 \let\babelFSfeatures\bbl@FSfeatures
4539 \babelFSfeatures}

```



```

4540 \def\bb1@FSfeatures#1#2{%
4541   \expandafter\addto\csname extras#1\endcsname{%
4542     \babel@save\bb1@langfeatures
4543     \edef\bb1@langfeatures{#2,}}
4544 \</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.

```

4545 \<{*Footnote changes}> ≡
4546 \bb1@trace{Bidi footnotes}
4547 \ifnum\bb1@bidimode>\z@
4548   \def\bb1@footnote#1#2#3{%
4549     \@ifnextchar[%
4550       {\bb1@footnote@o{#1}{#2}{#3}}%
4551       {\bb1@footnote@x{#1}{#2}{#3}}}
4552   \long\def\bb1@footnote@x#1#2#3#4{%
4553     \bgroup
4554     \select@language@x{\bb1@main@language}%
4555     \bb1@fn@footnote{#2#1{\ignorespaces#4}#3}%
4556     \egroup}
4557   \long\def\bb1@footnote@o#1#2#3[#4]#5{%
4558     \bgroup
4559     \select@language@x{\bb1@main@language}%
4560     \bb1@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
4561     \egroup}
4562   \def\bb1@footnotetext#1#2#3{%
4563     \@ifnextchar[%
4564       {\bb1@footnotetext@o{#1}{#2}{#3}}%
4565       {\bb1@footnotetext@x{#1}{#2}{#3}}}
4566   \long\def\bb1@footnotetext@x#1#2#3#4{%
4567     \bgroup
4568     \select@language@x{\bb1@main@language}%
4569     \bb1@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
4570     \egroup}
4571   \long\def\bb1@footnotetext@o#1#2#3[#4]#5{%
4572     \bgroup
4573     \select@language@x{\bb1@main@language}%
4574     \bb1@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
4575     \egroup}
4576   \def\BabelFootnote#1#2#3#4{%
4577     \ifx\bb1@fn@footnote\@undefined
4578       \let\bb1@fn@footnote\footnote
4579     \fi
4580     \ifx\bb1@fn@footnotetext\@undefined
4581       \let\bb1@fn@footnotetext\footnotetext
4582     \fi
4583     \bb1@ifblank{#2}%
4584       {\def#1{\bb1@footnote{\@firstofone}{#3}{#4}}
4585        \@namedef{\bb1@stripslash#1text}%
4586        {\bb1@footnotetext{\@firstofone}{#3}{#4}}}%
4587       {\def#1{\bb1@exp{\bb1@footnote{\@foreignlanguage{#2}}}{#3}{#4}}%
4588        \@namedef{\bb1@stripslash#1text}%
4589        {\bb1@exp{\bb1@footnotetext{\@foreignlanguage{#2}}}{#3}{#4}}}%

```

```

4590 \fi
4591 <</Footnote changes>>

```

Now, the code.

```

4592 (*xetex)
4593 \def\BabelStringsDefault{unicode}
4594 \let\xebbl@stop\relax
4595 \AddBabelHook{xetex}{encodedcommands}{%
4596   \def\bbl@tempa{#1}%
4597   \ifx\bbl@tempa\@empty
4598     \XeTeXinputencoding"bytes"%
4599   \else
4600     \XeTeXinputencoding"#1"%
4601   \fi
4602   \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
4603 \AddBabelHook{xetex}{stopcommands}{%
4604   \xebbl@stop
4605   \let\xebbl@stop\relax}
4606 \def\bbl@intraspace#1 #2 #3@@{%
4607   \bbl@csarg\gdef{\xeisp@\language}%
4608     {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
4609 \def\bbl@intrapenalty#1@@{%
4610   \bbl@csarg\gdef{\xeipn@\language}%
4611     {\XeTeXlinebreakpenalty #1\relax}}
4612 \def\bbl@provide@intraspace{%
4613   \bbl@xin@\bbl@cl{lnbrk}}{s}%
4614   \ifin\else\bbl@xin@\bbl@cl{lnbrk}}{c}\fi
4615   \ifin@
4616     \bbl@ifunset{\bbl@intsp@\language}{}%
4617     {\expandafter\ifx\csname bbl@intsp@\language\endcsname\@empty\else
4618       \ifx\bbl@KVP@intraspace\@nil
4619         \bbl@exp{%
4620           \\bbl@intraspace\bbl@cl{intsp}}\\@@}%
4621     \fi
4622     \ifx\bbl@KVP@intrapenalty\@nil
4623       \bbl@intrapenalty0@@
4624     \fi
4625   \fi
4626   \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
4627     \expandafter\bbl@intraspace\bbl@KVP@intraspace@@
4628   \fi
4629   \ifx\bbl@KVP@intrapenalty\@nil\else
4630     \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty@@
4631   \fi
4632   \bbl@exp{%
4633     \\bbl@add\<extras\language>{%
4634       \XeTeXlinebreaklocale "\bbl@cl{tbcpr}"%
4635       \<bbl@xeisp@\language>%
4636       \<bbl@xeipn@\language>%
4637       \\bbl@toglobal\<extras\language>%
4638       \\bbl@add\<noextras\language>{%
4639         \XeTeXlinebreaklocale "en"%
4640         \\bbl@toglobal\<noextras\language>}}%
4641   \ifx\bbl@ispacesize\undefined
4642     \gdef\bbl@ispacesize{\bbl@cl{xeisp}}%
4643     \ifx\AtBeginDocument\@notprerr
4644       \expandafter\@secondoftwo % to execute right now
4645     \fi
4646     \AtBeginDocument{%

```

```

4647         \expandafter\bb1@add
4648         \csname selectfont \endcsname{\bb1@ispace size}%
4649         \expandafter\bb1@tglobal\csname selectfont \endcsname}%
4650     \fi}%
4651 \fi}
4652 \ifx\DisableBabelHook\undefined\endinput\fi
4653 \AddBabelHook{babel-fontspec}{afterextras}{\bb1@switchfont}
4654 \AddBabelHook{babel-fontspec}{beforestart}{\bb1@ckeckstdfonts}
4655 \DisableBabelHook{babel-fontspec}
4656 <<Font selection>>
4657 \input txtbabel.def
4658 </xetex>

```

13.2 Layout

In progress.

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titles, and geometry.

\bb1@startskip and \bb1@endskip are available to package authors. Thanks to the T_EX expansion mechanism the following constructs are valid: \adim\bb1@startskip, \advance\bb1@startskip\adim, \bb1@startskip\adim.

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

```

4659 <*texxet>
4660 \providecommand\bb1@provide@intraspace{}
4661 \bb1@trace{Redefinitions for bidi layout}
4662 \def\bb1@sspre@caption{%
4663     \bb1@exp{\everyhbox{\bb1@textdir\bb1@cs{wdir\bb1@main@language}}}}
4664 \ifx\bb1@opt@layout\@nnil\endinput\fi % No layout
4665 \def\bb1@startskip{\ifcase\bb1@thepardir\leftskip\else\rightskip\fi}
4666 \def\bb1@endskip{\ifcase\bb1@thepardir\rightskip\else\leftskip\fi}
4667 \ifx\bb1@beforeforeign\leavevmode % A poor test for bidi=
4668     \def\hangfrom#1{%
4669         \setbox\@tempboxa\hbox{#1}%
4670         \hangindent\ifcase\bb1@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
4671         \noindent\box\@tempboxa}
4672 \def\raggedright{%
4673     \let\@centercr
4674     \bb1@startskip\z@skip
4675     \@rightskip\@flushglue
4676     \bb1@endskip\@rightskip
4677     \parindent\z@
4678     \parfillskip\bb1@startskip}
4679 \def\raggedleft{%
4680     \let\@centercr
4681     \bb1@startskip\@flushglue
4682     \bb1@endskip\z@skip
4683     \parindent\z@
4684     \parfillskip\bb1@endskip}
4685 \fi
4686 \IfBabelLayout{lists}
4687 {\bb1@sreplace\list
4688     {\@totalleftmargin\leftmargin}{\@totalleftmargin\bb1@listleftmargin}%
4689     \def\bb1@listleftmargin{%
4690         \ifcase\bb1@thepardir\leftmargin\else\rightmargin\fi}%
4691     \ifcase\bb1@engine
4692         \def\labelenumii{}\theenumii{}}% pdfTeX doesn't reverse ()

```

```

4693 \def\p@enumiii{\p@enumii}\theenumii}%
4694 \fi
4695 \bbl@sreplace\@verbatim
4696 {\leftskip\@totalleftmargin}%
4697 {\bbl@startskip\textwidth
4698 \advance\bbl@startskip-\linewidth}%
4699 \bbl@sreplace\@verbatim
4700 {\rightskip\z@skip}%
4701 {\bbl@endskip\z@skip}}%
4702 {}
4703 \IfBabelLayout{contents}
4704 {\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}%
4705 \bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}
4706 {}
4707 \IfBabelLayout{columns}
4708 {\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputbox}%
4709 \def\bbl@outputbox#1{%
4710 \hb@xt@\textwidth{%
4711 \hskip\columnwidth
4712 \hfil
4713 {\normalcolor\vrule \@width\columnseprule}%
4714 \hfil
4715 \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
4716 \hskip-\textwidth
4717 \hb@xt@\columnwidth{\box\@outputbox \hss}%
4718 \hskip\columnsep
4719 \hskip\columnwidth}}}%
4720 {}
4721 <<Footnote changes>>
4722 \IfBabelLayout{footnotes}%
4723 {\BabelFootnote\footnote\language\{}}%
4724 \BabelFootnote\localfootnote\language\{}}%
4725 \BabelFootnote\mainfootnote\{}}{}%
4726 {}
4727 \IfBabelLayout{counters}%
4728 {\let\bbl@latinarabic=\@arabic
4729 \def\@arabic#1{\bbl@sublr{\bbl@latinarabic#1}}%
4730 \let\bbl@asciroman=\@roman
4731 \def\@roman#1{\bbl@sublr{\ensureascii{\bbl@asciroman#1}}}%
4732 \let\bbl@asciiRoman=\@Roman
4733 \def\@Roman#1{\bbl@sublr{\ensureascii{\bbl@asciiRoman#1}}}}{}
4734 </texxet>

```

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

13.3 LuaTeX

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

The names \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`). FIX - This isn’t true anymore. For the moment, a dangerous approach is used - just allocate a high random number and cross the fingers. To complicate things, `etex.sty` changes the way languages are allocated.

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

```

4735 (*luatex)
4736 \ifx\AddBabelHook\undefined % When plain.def, babel.sty starts
4737 \bbl@trace{Read language.dat}
4738 \ifx\bbl@readstream\undefined
4739   \csname newread\endcsname\bbl@readstream
4740 \fi
4741 \begingroup
4742   \toks@{}
4743   \count@ \z@ % 0=start, 1=0th, 2=normal
4744   \def\bbl@process@line#1#2 #3 #4 {%
4745     \ifx=#1%
4746       \bbl@process@synonym{#2}%
4747     \else
4748       \bbl@process@language{#1#2}{#3}{#4}%
4749     \fi
4750     \ignorespaces}
4751   \def\bbl@manylang{%
4752     \ifnum\bbl@last>\@ne
4753       \bbl@info{Non-standard hyphenation setup}%
4754     \fi
4755     \let\bbl@manylang\relax}
4756   \def\bbl@process@language#1#2#3{%
4757     \ifcase\count@
4758       \@ifundefined{zth#1}{\count@ \tw@}{\count@ \@ne}%
4759     \or
4760       \count@ \tw@
4761     \fi
4762     \ifnum\count@=\tw@
4763       \expandafter\addlanguage\csname l@#1\endcsname
4764       \language\allocationnumber
4765     \chardef\bbl@last\allocationnumber

```

```

4766 \bbl@manylang
4767 \let\bbl@elt\relax
4768 \xdef\bbl@languages{%
4769 \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
4770 \fi
4771 \the\toks@
4772 \toks@{}}
4773 \def\bbl@process@synonym@aux#1#2{%
4774 \global\expandafter\chardef\csname l@#1\endcsname#2\relax
4775 \let\bbl@elt\relax
4776 \xdef\bbl@languages{%
4777 \bbl@languages\bbl@elt{#1}{#2}{}}}%
4778 \def\bbl@process@synonym#1{%
4779 \ifcase\count@
4780 \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
4781 \or
4782 \@ifundefined{zth@#1}{\bbl@process@synonym@aux{#1}{0}}}%
4783 \else
4784 \bbl@process@synonym@aux{#1}{\the\bbl@last}%
4785 \fi}
4786 \ifx\bbl@languages@\undefined % Just a (sensible?) guess
4787 \chardef\l@english\z@
4788 \chardef\l@USenglish\z@
4789 \chardef\bbl@last\z@
4790 \global\@namedef{\bbl@hyphendata@0}{\hyphen.tex}}
4791 \gdef\bbl@languages{%
4792 \bbl@elt{english}{0}{\hyphen.tex}}%
4793 \bbl@elt{USenglish}{0}{}}
4794 \else
4795 \global\let\bbl@languages@format\bbl@languages
4796 \def\bbl@elt#1#2#3#4{% Remove all except language 0
4797 \ifnum#2>\z@\else
4798 \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
4799 \fi}%
4800 \xdef\bbl@languages{\bbl@languages}%
4801 \fi
4802 \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}} % Define flags
4803 \bbl@languages
4804 \openin\bbl@readstream=language.dat
4805 \ifeof\bbl@readstream
4806 \bbl@warning{I couldn't find language.dat. No additional\\%
4807 patterns loaded. Reported}%
4808 \else
4809 \loop
4810 \endlinechar\m@ne
4811 \read\bbl@readstream to \bbl@line
4812 \endlinechar`\^^M
4813 \if T\ifeof\bbl@readstream F\fi T\relax
4814 \ifx\bbl@line\empty\else
4815 \edef\bbl@line{\bbl@line\space\space\space}%
4816 \expandafter\bbl@process@line\bbl@line\relax
4817 \fi
4818 \repeat
4819 \fi
4820 \endgroup
4821 \bbl@trace{Macros for reading patterns files}
4822 \def\bbl@get@enc#1:#2:#3\@@{\def\bbl@hyph@enc{#2}}
4823 \ifx\babelcatcodetablenum\undefined
4824 \ifx\newcatcodetable\undefined

```

```

4825 \def\babelcatcodetablenum{5211}
4826 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4827 \else
4828 \newcatcodetable\babelcatcodetablenum
4829 \newcatcodetable\bbl@pattcodes
4830 \fi
4831 \else
4832 \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4833 \fi
4834 \def\bbl@luapatterns#1#2{%
4835 \bbl@get@enc#1::@@@
4836 \setbox\z@\hbox\bgroup
4837 \begingroup
4838 \savecatcodetable\babelcatcodetablenum\relax
4839 \initcatcodetable\bbl@pattcodes\relax
4840 \catcodetable\bbl@pattcodes\relax
4841 \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
4842 \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\-=13
4843 \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
4844 \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
4845 \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
4846 \catcode`\`=12 \catcode`\'=12 \catcode`\`=12
4847 \input #1\relax
4848 \catcodetable\babelcatcodetablenum\relax
4849 \endgroup
4850 \def\bbl@tempa{#2}%
4851 \ifx\bbl@tempa@empty\else
4852 \input #2\relax
4853 \fi
4854 \egroup}%
4855 \def\bbl@patterns@lua#1{%
4856 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
4857 \csname l@#1\endcsname
4858 \edef\bbl@tempa{#1}%
4859 \else
4860 \csname l@#1:\f@encoding\endcsname
4861 \edef\bbl@tempa{#1:\f@encoding}%
4862 \fi\relax
4863 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
4864 \@ifundefined{bbl@hyphendata@the\language}%
4865 {\def\bbl@elt##1##2##3##4{%
4866 \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
4867 \def\bbl@tempb{##3}%
4868 \ifx\bbl@tempb@empty\else % if not a synonymous
4869 \def\bbl@tempc{##3}{##4}%
4870 \fi
4871 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4872 \fi}%
4873 \bbl@languages
4874 \@ifundefined{bbl@hyphendata@the\language}%
4875 {\bbl@info{No hyphenation patterns were set for\the
4876 language '\bbl@tempa'. Reported}}%
4877 {\expandafter\expandafter\expandafter\bbl@luapatterns
4878 \csname bbl@hyphendata@the\language\endcsname}}}%
4879 \endinput\fi
4880 % Here ends \ifx\AddBabelHook\@undefined
4881 % A few lines are only read by hyphen.cfg
4882 \ifx\DisableBabelHook\@undefined
4883 \AddBabelHook{luatex}{everylanguage}{%

```

```

4884 \def\process@language##1##2##3{%
4885 \def\process@line####1####2 ####3 ####4 {}}
4886 \AddBabelHook{luatex}{loadpatterns}{%
4887 \input #1\relax
4888 \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
4889 {#{1}{}}}
4890 \AddBabelHook{luatex}{loadexceptions}{%
4891 \input #1\relax
4892 \def\bbl@tempb##1##2{#{1}{#1}}%
4893 \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
4894 {\expandafter\expandafter\expandafter\bbl@tempb
4895 \csname bbl@hyphendata@the\language\endcsname}}
4896 \endinput\fi
4897 % Here stops reading code for hyphen.cfg
4898 % The following is read the 2nd time it's loaded
4899 \begingroup % TODO - to a lua file
4900 \catcode`\%=12
4901 \catcode`\'=12
4902 \catcode`\%=12
4903 \catcode`\:=12
4904 \directlua{
4905 Babel = Babel or {}
4906 function Babel.bytes(line)
4907 return line:gsub(".",
4908 function (chr) return unicode.utf8.char(string.byte(chr)) end)
4909 end
4910 function Babel.begin_process_input()
4911 if luatexbase and luatexbase.add_to_callback then
4912 luatexbase.add_to_callback('process_input_buffer',
4913 Babel.bytes,'Babel.bytes')
4914 else
4915 Babel.callback = callback.find('process_input_buffer')
4916 callback.register('process_input_buffer',Babel.bytes)
4917 end
4918 end
4919 function Babel.end_process_input ()
4920 if luatexbase and luatexbase.remove_from_callback then
4921 luatexbase.remove_from_callback('process_input_buffer','Babel.bytes')
4922 else
4923 callback.register('process_input_buffer',Babel.callback)
4924 end
4925 end
4926 function Babel.addpatterns(pp, lg)
4927 local lg = lang.new(lg)
4928 local pats = lang.patterns(lg) or ''
4929 lang.clear_patterns(lg)
4930 for p in pp:gmatch('[^%s]+') do
4931 ss = ''
4932 for i in string.utfcharacters(p:gsub('%d', '')) do
4933 ss = ss .. '%d?' .. i
4934 end
4935 ss = ss:gsub('^%%d%?%', '%%.') .. '%d?'
4936 ss = ss:gsub('%.%d%?$', '%%.')
4937 pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
4938 if n == 0 then
4939 tex.sprint(
4940 [[\string\csname\space bbl@info\endcsname{New pattern: }
4941 .. p .. [{}]])
4942 pats = pats .. ' ' .. p

```



```

4943     else
4944         tex.sprint(
4945             [[\string\csname\space bbl@info\endcsname{Renew pattern: }]
4946             .. p .. [{}]])
4947     end
4948 end
4949 lang.patterns(lg, pats)
4950 end
4951 }
4952 \endgroup
4953 \ifx\newattribute\@undefined\else
4954 \newattribute\bbl@attr@locale
4955 \directlua{ Babel.attr_locale = luatexbase.registernumber'bbl@attr@locale'}
4956 \AddBabelHook{luatex}{beforeextras}{%
4957     \setattribute\bbl@attr@locale\localeid}
4958 \fi
4959 \def\BabelStringsDefault{unicode}
4960 \let\luabbl@stop\relax
4961 \AddBabelHook{luatex}{encodedcommands}{%
4962     \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
4963     \ifx\bbl@tempa\bbl@tempb\else
4964         \directlua{Babel.begin_process_input()}%
4965         \def\luabbl@stop{%
4966             \directlua{Babel.end_process_input()}}%
4967     \fi}%
4968 \AddBabelHook{luatex}{stopcommands}{%
4969     \luabbl@stop
4970     \let\luabbl@stop\relax}
4971 \AddBabelHook{luatex}{patterns}{%
4972     \@ifundefined{bbl@hyphendata@the\language}%
4973     {\def\bbl@elt##1##2##3##4{%
4974         \ifnum##2=\csname l@#2\endcsname % #2=spanish, dutch:OT1...
4975         \def\bbl@tempb{##3}%
4976         \ifx\bbl@tempb\@empty\else % if not a synonymous
4977             \def\bbl@tempc{##3}{##4}}%
4978         \fi
4979         \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4980         \fi}%
4981     \bbl@languages
4982     \@ifundefined{bbl@hyphendata@the\language}%
4983     {\bbl@info{No hyphenation patterns were set for\%
4984         language '#2'. Reported}}%
4985     {\expandafter\expandafter\expandafter\bbl@luapatterns
4986         \csname bbl@hyphendata@the\language\endcsname}}}%
4987 \@ifundefined{bbl@patterns@}{}%
4988 \begingroup
4989 \bbl@xin@{\, \number\language,}{, \bbl@pttnlist}%
4990 \ifin@ \else
4991     \ifx\bbl@patterns@\@empty \else
4992         \directlua{ Babel.addpatterns(
4993             [[\bbl@patterns@]], \number\language) }%
4994     \fi
4995     \@ifundefined{bbl@patterns@#1}%
4996     \@empty
4997     {\directlua{ Babel.addpatterns(
4998         [[\space\csname bbl@patterns@#1\endcsname]],
4999         \number\language) }}%
5000     \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
5001 \fi

```

```

5002 \endgroup}%
5003 \bbl@exp{%
5004 \bbl@ifunset{bbl@prehc@<language>}{}%
5005 {\bbl@ifblank{bbl@cs{prehc@<language>}}{}}%
5006 {\prehyphenchar=bbl@c1{prehc}\relax}}

```

`\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.

```

5007 \@onlypreamble\babelpatterns
5008 \AtEndOfPackage{%
5009 \newcommand\babelpatterns[2][\@empty]{%
5010 \ifx\bbl@patterns@<relax>
5011 \let\bbl@patterns@<empty>
5012 \fi
5013 \ifx\bbl@pttnlist<empty>\else
5014 \bbl@warning{%
5015 You must not intermingle \string\selectlanguage\space and\%
5016 \string\babelpatterns\space or some patterns will not\%
5017 be taken into account. Reported}%
5018 \fi
5019 \ifx<empty>#1%
5020 \protected@edef\bbl@patterns@{\bbl@patterns@<space>#2}%
5021 \else
5022 \edef\bbl@tempb{\zap<space>#1 <empty>}%
5023 \bbl@for\bbl@tempa\bbl@tempb{%
5024 \bbl@fixname\bbl@tempa
5025 \bbl@iflanguage\bbl@tempa{%
5026 \bbl@csarg\protected@edef{patterns@<bbl@tempa>}{%
5027 \ifundefined{bbl@patterns@<bbl@tempa>}%
5028 <empty>
5029 {\csname bbl@patterns@<bbl@tempa>\endcsname\space}%
5030 #2}}}%
5031 \fi}}

```

13.4 Southeast Asian scripts

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

```

5032 % TODO - to a lua file
5033 \directlua{
5034 Babel = Babel or {}
5035 Babel.linebreaking = Babel.linebreaking or {}
5036 Babel.linebreaking.before = {}
5037 Babel.linebreaking.after = {}
5038 Babel.locale = {} % Free to use, indexed with \localeid
5039 function Babel.linebreaking.add_before(func)
5040 tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5041 table.insert(Babel.linebreaking.before, func)
5042 end
5043 function Babel.linebreaking.add_after(func)
5044 tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5045 table.insert(Babel.linebreaking.after, func)
5046 end
5047 }

```

```

5048 \def\bbl@intraspace#1 #2 #3\@@{%
5049 \directlua{
5050   Babel = Babel or {}
5051   Babel.intraspaces = Babel.intraspaces or {}
5052   Babel.intraspaces['\csname bbl@sbc@language\endcsname'] = %
5053     {b = #1, p = #2, m = #3}
5054   Babel.locale_props[\the\localeid].intraspace = %
5055     {b = #1, p = #2, m = #3}
5056 }}
5057 \def\bbl@intrapenalty#1\@@{%
5058 \directlua{
5059   Babel = Babel or {}
5060   Babel.intrapenalties = Babel.intrapenalties or {}
5061   Babel.intrapenalties['\csname bbl@sbc@language\endcsname'] = #1
5062   Babel.locale_props[\the\localeid].intrapenalty = #1
5063 }}
5064 \begingroup
5065 \catcode`\%=12
5066 \catcode`\^=14
5067 \catcode`\'=12
5068 \catcode`\~=12
5069 \gdef\bbl@seaintraspace{^
5070 \let\bbl@seaintraspace\relax
5071 \directlua{
5072   Babel = Babel or {}
5073   Babel.sea_enabled = true
5074   Babel.sea_ranges = Babel.sea_ranges or {}
5075   function Babel.set_chrngs (script, chrng)
5076     local c = 0
5077     for s, e in string.gmatch(chrng..' ', '(.-%.%.(-)%s') do
5078       Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
5079       c = c + 1
5080     end
5081   end
5082   function Babel.sea_disc_to_space (head)
5083     local sea_ranges = Babel.sea_ranges
5084     local last_char = nil
5085     local quad = 655360 ^% 10 pt = 655360 = 10 * 65536
5086     for item in node.traverse(head) do
5087       local i = item.id
5088       if i == node.id'glyph' then
5089         last_char = item
5090       elseif i == 7 and item.subtype == 3 and last_char
5091         and last_char.char > 0x0C99 then
5092         quad = font.getfont(last_char.font).size
5093         for lg, rg in pairs(sea_ranges) do
5094           if last_char.char > rg[1] and last_char.char < rg[2] then
5095             lg = lg:sub(1, 4) ^% Remove trailing number of, eg, Cyril1
5096             local intraspace = Babel.intraspaces[lg]
5097             local intrapenalty = Babel.intrapenalties[lg]
5098             local n
5099             if intrapenalty ~= 0 then
5100               n = node.new(14, 0) ^% penalty
5101               n.penalty = intrapenalty
5102               node.insert_before(head, item, n)
5103             end
5104             n = node.new(12, 13) ^% (glue, spaceskip)
5105             node.setglue(n, intraspace.b * quad,
5106               intraspace.p * quad,

```

```

5107             intraspace.m * quad)
5108         node.insert_before(head, item, n)
5109         node.remove(head, item)
5110     end
5111 end
5112 end
5113 end
5114 end
5115 }^^
5116 \bbl@luahyphenate}
5117 \catcode`\%=14
5118 \gdef\bbl@cjkintraspacespace{%
5119 \let\bbl@cjkintraspacespace\relax
5120 \directlua{
5121     Babel = Babel or {}
5122     require'babel-data-cjk.lua'
5123     Babel.cjk_enabled = true
5124     function Babel.cjk_linebreak(head)
5125         local GLYPH = node.id'glyph'
5126         local last_char = nil
5127         local quad = 655360      % 10 pt = 655360 = 10 * 65536
5128         local last_class = nil
5129         local last_lang = nil
5130
5131         for item in node.traverse(head) do
5132             if item.id == GLYPH then
5133
5134                 local lang = item.lang
5135
5136                 local LOCALE = node.get_attribute(item,
5137                     luatexbase.registernumber'bbl@attr@locale')
5138                 local props = Babel.locale_props[LOCALE]
5139
5140                 local class = Babel.cjk_class[item.char].c
5141
5142                 if class == 'cp' then class = 'cl' end % )] as CL
5143                 if class == 'id' then class = 'I' end
5144
5145                 local br = 0
5146                 if class and last_class and Babel.cjk_breaks[last_class][class] then
5147                     br = Babel.cjk_breaks[last_class][class]
5148                 end
5149
5150                 if br == 1 and props.linebreak == 'c' and
5151                     lang ~= \the\l@nohyphenation\space and
5152                     last_lang ~= \the\l@nohyphenation then
5153                     local intrapenalty = props.intrapenalty
5154                     if intrapenalty ~= 0 then
5155                         local n = node.new(14, 0)      % penalty
5156                         n.penalty = intrapenalty
5157                         node.insert_before(head, item, n)
5158                     end
5159                     local intraspace = props.intraspace
5160                     local n = node.new(12, 13)          % (glue, spaceskip)
5161                     node.setglue(n, intraspace.b * quad,
5162                         intraspace.p * quad,
5163                         intraspace.m * quad)
5164                     node.insert_before(head, item, n)
5165                 end

```

```

5166
5167         if font.getfont(item.font) then
5168             quad = font.getfont(item.font).size
5169         end
5170         last_class = class
5171         last_lang = lang
5172     else % if penalty, glue or anything else
5173         last_class = nil
5174     end
5175 end
5176 lang.hyphenate(head)
5177 end
5178 }%
5179 \bbl@luahyphenate}
5180 \gdef\bbl@luahyphenate{%
5181 \let\bbl@luahyphenate\relax
5182 \directlua{
5183     luatexbase.add_to_callback('hyphenate',
5184     function (head, tail)
5185         if Babel.linebreaking.before then
5186             for k, func in ipairs(Babel.linebreaking.before) do
5187                 func(head)
5188             end
5189         end
5190         if Babel.cjk_enabled then
5191             Babel.cjk_linebreak(head)
5192         end
5193         lang.hyphenate(head)
5194         if Babel.linebreaking.after then
5195             for k, func in ipairs(Babel.linebreaking.after) do
5196                 func(head)
5197             end
5198         end
5199         if Babel.sea_enabled then
5200             Babel.sea_disc_to_space(head)
5201         end
5202     end,
5203     'Babel.hyphenate')
5204 }
5205 }
5206 \endgroup
5207 \def\bbl@provide@intraspace{%
5208 \bbl@ifunset{\bbl@intsp@{language}}{%
5209     {\expandafter\ifx\csname bbl@intsp@{language}\endcsname\@empty\else
5210         \bbl@xin@{\bbl@cl{lnbrk}}{c}%
5211         \ifin@ % cjk
5212         \bbl@cjk_intraspace
5213         \directlua{
5214             Babel = Babel or {}
5215             Babel.locale_props = Babel.locale_props or {}
5216             Babel.locale_props[\the\localeid].linebreak = 'c'
5217         }%
5218         \bbl@exp{\bbl@intraspace\bbl@cl{intsp}}{\bbl@intsp}%
5219         \ifx\bbl@KVP@intrapenalty\@nil
5220             \bbl@intrapenalty0\@
5221         \fi
5222     \else % sea
5223         \bbl@seaintraspace
5224         \bbl@exp{\bbl@intraspace\bbl@cl{intsp}}{\bbl@intsp}%

```

```

5225 \directlua{
5226     Babel = Babel or {}
5227     Babel.sea_ranges = Babel.sea_ranges or {}
5228     Babel.set_chranges('\bbl@c1{sbc}p',
5229                        '\bbl@c1{chrng}')
5230 }%
5231 \ifx\bbl@KVP@intrapenalty\@nil
5232     \bbl@intrapenalty0\@@
5233 \fi
5234 \fi
5235 \fi
5236 \ifx\bbl@KVP@intrapenalty\@nil\else
5237     \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
5238 \fi}}

```

13.5 CJK line breaking

Minimal line breaking for CJK scripts, mainly intended for simple documents and short texts as a secondary language. Only line breaking, with a little stretching for justification, without any attempt to adjust the spacing. It is based on (but does not strictly follow) the Unicode algorithm.

We first need a little table with the corresponding line breaking properties. A few characters have an additional key for the width (fullwidth vs. halfwidth), not yet used. There is a separate file, defined below.

Work in progress.

Common stuff.

```

5239 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
5240 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@cckstdfont}
5241 \DisableBabelHook{babel-fontspec}
5242 <<Font selection>>

```

13.6 Automatic fonts and ids switching

After defining the blocks for a number of scripts (must be extended and very likely fine tuned), we define a short function which just traverse the node list to carry out the replacements. The table `loc_to_scr` gets the locale from a script range (note the locale is the key, and that there is an intermediate table built on the fly for optimization). This locale is then used to get the `\language` and the `\localeid` as stored in `locale_props`, as well as the font (as requested). In the latter table a key starting with `/` maps the font from the global one (the key) to the local one (the value). Maths are skipped and discretionaries are handled in a special way.

```

5243% TODO - to a lua file
5244 \directlua{
5245 Babel.script_blocks = {
5246   ['dflt'] = {},
5247   ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
5248              {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
5249   ['Armn'] = {{0x0530, 0x058F}},
5250   ['Beng'] = {{0x0980, 0x09FF}},
5251   ['Cher'] = {{0x13A0, 0x13FF}, {0xAB70, 0xABBF}},
5252   ['Copt'] = {{0x03E2, 0x03EF}, {0x2C80, 0x2CFF}, {0x102E0, 0x102FF}},
5253   ['Cyr1'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
5254              {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
5255   ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
5256   ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
5257              {0xAB00, 0xAB2F}},

```

```

5258 ['Geor'] = {{0x10A0, 0x10FF}, {0x2D00, 0x2D2F}},
5259 % Don't follow strictly Unicode, which places some Coptic letters in
5260 % the 'Greek and Coptic' block
5261 ['Grek'] = {{0x0370, 0x03E1}, {0x03F0, 0x03FF}, {0x1F00, 0x1FFF}},
5262 ['Hans'] = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
5263             {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
5264             {0xF900, 0xFAFF}, {0xFE30, 0xFE4F}, {0xFF00, 0xFFEF},
5265             {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
5266             {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
5267             {0x2CEB0, 0x2EBEF}, {0x2F800, 0x2FA1F}},
5268 ['Hebr'] = {{0x0590, 0x05FF}},
5269 ['Jpan'] = {{0x3000, 0x303F}, {0x3040, 0x309F}, {0x30A0, 0x30FF},
5270             {0x4E00, 0x9FAF}, {0xFF00, 0xFFEF}},
5271 ['Khmr'] = {{0x1780, 0x17FF}, {0x19E0, 0x19FF}},
5272 ['Knda'] = {{0x0C80, 0x0CFF}},
5273 ['Kore'] = {{0x1100, 0x11FF}, {0x3000, 0x303F}, {0x3130, 0x318F},
5274             {0x4E00, 0x9FAF}, {0xA960, 0xA97F}, {0xAC00, 0xD7AF},
5275             {0xD7B0, 0xD7FF}, {0xFF00, 0xFFEF}},
5276 ['Laoo'] = {{0x0E80, 0x0EFF}},
5277 ['Latn'] = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
5278             {0x0180, 0x024F}, {0x1E00, 0x1EFF}, {0x2C60, 0x2C7F},
5279             {0xA720, 0xA7FF}, {0xAB30, 0xAB6F}},
5280 ['Mahj'] = {{0x11150, 0x1117F}},
5281 ['Mlym'] = {{0x0D00, 0x0D7F}},
5282 ['Mymr'] = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA9E0, 0xA9FF}},
5283 ['Orya'] = {{0x0B00, 0x0B7F}},
5284 ['Sinh'] = {{0x0D80, 0x0DFF}, {0x111E0, 0x111FF}},
5285 ['Sycr'] = {{0x0700, 0x074F}, {0x0860, 0x086F}},
5286 ['Taml'] = {{0x0B80, 0x0BFF}},
5287 ['Telu'] = {{0x0C00, 0x0C7F}},
5288 ['Tfng'] = {{0x2D30, 0x2D7F}},
5289 ['Thai'] = {{0x0E00, 0x0E7F}},
5290 ['Tibt'] = {{0x0F00, 0x0FFF}},
5291 ['Vaii'] = {{0xA500, 0xA63F}},
5292 ['Yiii'] = {{0xA000, 0xA48F}, {0xA490, 0xA4CF}}
5293 }
5294
5295 Babel.script_blocks.Cyrs = Babel.script_blocks.Cyrl
5296 Babel.script_blocks.Hant = Babel.script_blocks.Hans
5297 Babel.script_blocks.Kana = Babel.script_blocks.Jpan
5298
5299 function Babel.locale_map(head)
5300   if not Babel.locale_mapped then return head end
5301
5302   local LOCALE = luatexbase.registernumber'bbl@attr@locale'
5303   local GLYPH = node.id('glyph')
5304   local inmath = false
5305   local toloc_save
5306   for item in node.traverse(head) do
5307     local toloc
5308     if not inmath and item.id == GLYPH then
5309       % Optimization: build a table with the chars found
5310       if Babel.chr_to_loc[item.char] then
5311         toloc = Babel.chr_to_loc[item.char]
5312       else
5313         for lc, maps in pairs(Babel.loc_to_scr) do
5314           for _, rg in pairs(maps) do
5315             if item.char >= rg[1] and item.char <= rg[2] then
5316               Babel.chr_to_loc[item.char] = lc

```

```

5317         toloc = lc
5318         break
5319     end
5320 end
5321 end
5322 end
5323 % Now, take action, but treat composite chars in a different
5324 % fashion, because they 'inherit' the previous locale. Not yet
5325 % optimized.
5326 if not toloc and
5327     (item.char >= 0x0300 and item.char <= 0x036F) or
5328     (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
5329     (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
5330     toloc = toloc_save
5331 end
5332 if toloc and toloc > -1 then
5333     if Babel.locale_props[toloc].lg then
5334         item.lang = Babel.locale_props[toloc].lg
5335         node.set_attribute(item, LOCALE, toloc)
5336     end
5337     if Babel.locale_props[toloc]['/'..item.font] then
5338         item.font = Babel.locale_props[toloc]['/'..item.font]
5339     end
5340     toloc_save = toloc
5341 end
5342 elseif not inmath and item.id == 7 then
5343     item.replace = item.replace and Babel.locale_map(item.replace)
5344     item.pre      = item.pre and Babel.locale_map(item.pre)
5345     item.post     = item.post and Babel.locale_map(item.post)
5346 elseif item.id == node.id'math' then
5347     inmath = (item.subtype == 0)
5348 end
5349 end
5350 return head
5351 end
5352 }

```

The code for `\babelcharproperty` is straightforward. Just note the modified lua table can be different.

```

5353 \newcommand\babelcharproperty[1]{%
5354   \count@=#1\relax
5355   \ifvmode
5356     \expandafter\bbl@chprop
5357   \else
5358     \bbl@error{\string\babelcharproperty\space can be used only in\\%
5359               vertical mode (preamble or between paragraphs)}%
5360     {See the manual for futher info}%
5361   \fi}
5362 \newcommand\bbl@chprop[3][\the\count@]{%
5363   \@tempcnta=#1\relax
5364   \bbl@ifunset{\bbl@chprop@#2}%
5365   {\bbl@error{No property named '#2'. Allowed values are\\%
5366             direction (bc), mirror (bmg), and linebreak (lb)}%
5367    {See the manual for futher info}}%
5368   }%
5369   \loop
5370     \bbl@cs{chprop@#2}{#3}%
5371   \ifnum\count@<\@tempcnta
5372     \advance\count@\@ne

```



```

5373 \repeat}
5374 \def\bbl@chprop@direction#1{%
5375 \directlua{
5376   Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5377   Babel.characters[\the\count@]['d'] = '#1'
5378 }}
5379 \let\bbl@chprop@bc\bbl@chprop@direction
5380 \def\bbl@chprop@mirror#1{%
5381 \directlua{
5382   Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5383   Babel.characters[\the\count@]['m'] = '\number#1'
5384 }}
5385 \let\bbl@chprop@bmg\bbl@chprop@mirror
5386 \def\bbl@chprop@linebreak#1{%
5387 \directlua{
5388   Babel.cjk_characters[\the\count@] = Babel.cjk_characters[\the\count@] or {}
5389   Babel.cjk_characters[\the\count@]['c'] = '#1'
5390 }}
5391 \let\bbl@chprop@lb\bbl@chprop@linebreak
5392 \def\bbl@chprop@locale#1{%
5393 \directlua{
5394   Babel.chr_to_loc = Babel.chr_to_loc or {}
5395   Babel.chr_to_loc[\the\count@] =
5396     \bbl@ifblank{#1}{-1000}{\the\bbl@cs{id@#1}}\space
5397 }}

```

Post-handling hyphenation patterns for non-standard rules, like ff to ff-f. There are still some issues with speed (not very slow, but still slow).

After declaring the table containing the patterns with their replacements, we define some auxiliary functions: `str_to_nodes` converts the string returned by a function to a node list, taking the node at base as a model (font, language, etc.); `fetch_word` fetches a series of glyphs and discretionaries, which pattern is matched against (if there is a match, it is called again before trying other patterns, and this is very likely the main bottleneck). `post_hyphenate_replace` is the callback applied after `lang.hyphenate`. This means the automatic hyphenation points are known. As empty captures return a byte position (as explained in the `luatex` manual), we must convert it to a utf8 position. With `first`, the last byte can be the leading byte in a utf8 sequence, so we just remove it and add 1 to the resulting length. With `last` we must take into account the capture position points to the next character. Here `word_head` points to the starting node of the text to be matched.

```

5398 \begingroup % TODO - to a lua file
5399 \catcode`\~=12
5400 \catcode`\#=12
5401 \catcode`\%=12
5402 \catcode`\&=14
5403 \directlua{
5404   Babel.linebreaking.replacements = {}
5405   Babel.linebreaking.replacements[0] = {} &% pre
5406   Babel.linebreaking.replacements[1] = {} &% post
5407
5408   &% Discretionaries contain strings as nodes
5409   function Babel.str_to_nodes(fn, matches, base)
5410     local n, head, last
5411     if fn == nil then return nil end
5412     for s in string.utfvalues(fn(matches)) do
5413       if base.id == 7 then
5414         base = base.replace
5415       end
5416       n = node.copy(base)

```

```

5417     n.char    = s
5418     if not head then
5419         head = n
5420     else
5421         last.next = n
5422     end
5423     last = n
5424 end
5425 return head
5426 end
5427
5428 Babel.fetch_subtext = {}
5429
5430 %% Merging both functions doesn't seem feasible, because there are too
5431 %% many differences.
5432 Babel.fetch_subtext[0] = function(head)
5433     local word_string = ''
5434     local word_nodes = {}
5435     local lang
5436     local item = head
5437     local inmath = false
5438
5439     while item do
5440
5441         if item.id == 11 then
5442             inmath = (item.subtype == 0)
5443         end
5444
5445         if inmath then
5446             %% pass
5447
5448         elseif item.id == 29 then
5449             local locale = node.get_attribute(item, Babel.attr_locale)
5450
5451             if lang == locale or lang == nil then
5452                 if (item.char ~= 124) then %% ie, not | = space
5453                     lang = lang or locale
5454                     word_string = word_string .. unicode.utf8.char(item.char)
5455                     word_nodes[#word_nodes+1] = item
5456                 end
5457             else
5458                 break
5459             end
5460
5461         elseif item.id == 12 and item.subtype == 13 then
5462             word_string = word_string .. '|'
5463             word_nodes[#word_nodes+1] = item
5464
5465             %% Ignore leading unrecognized nodes, too.
5466         elseif word_string ~= '' then
5467             word_string = word_string .. Babel.us_char
5468             word_nodes[#word_nodes+1] = item %% Will be ignored
5469         end
5470
5471         item = item.next
5472     end
5473
5474     %% Here and above we remove some trailing chars but not the
5475     %% corresponding nodes. But they aren't accessed.

```

```

5476     if word_string:sub(-1) == '|' then
5477         word_string = word_string:sub(1,-2)
5478     end
5479     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5480     return word_string, word_nodes, item, lang
5481 end
5482
5483 Babel.fetch_subtext[1] = function(head)
5484     local word_string = ''
5485     local word_nodes = {}
5486     local lang
5487     local item = head
5488     local inmath = false
5489
5490     while item do
5491
5492         if item.id == 11 then
5493             inmath = (item.subtype == 0)
5494         end
5495
5496         if inmath then
5497             && pass
5498
5499         elseif item.id == 29 then
5500             if item.lang == lang or lang == nil then
5501                 if (item.char ~= 124) and (item.char ~= 61) then && not =, not |
5502                     lang = lang or item.lang
5503                     word_string = word_string .. unicode.utf8.char(item.char)
5504                     word_nodes[#word_nodes+1] = item
5505                 end
5506             else
5507                 break
5508             end
5509
5510         elseif item.id == 7 and item.subtype == 2 then
5511             word_string = word_string .. '='
5512             word_nodes[#word_nodes+1] = item
5513
5514         elseif item.id == 7 and item.subtype == 3 then
5515             word_string = word_string .. '|'
5516             word_nodes[#word_nodes+1] = item
5517
5518             && (1) Go to next word if nothing was found, and (2) implicitly
5519             && remove leading USs.
5520         elseif word_string == '' then
5521             && pass
5522
5523             && This is the responsible for splitting by words.
5524         elseif (item.id == 12 and item.subtype == 13) then
5525             break
5526
5527         else
5528             word_string = word_string .. Babel.us_char
5529             word_nodes[#word_nodes+1] = item && Will be ignored
5530         end
5531
5532         item = item.next
5533     end
5534

```

```

5535     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5536     return word_string, word_nodes, item, lang
5537 end
5538
5539 function Babel.pre_hyphenate_replace(head)
5540     Babel.hyphenate_replace(head, 0)
5541 end
5542
5543 function Babel.post_hyphenate_replace(head)
5544     Babel.hyphenate_replace(head, 1)
5545 end
5546
5547 Babel.us_char = string.char(31)
5548
5549 function Babel.hyphenate_replace(head, mode)
5550     local u = unicode.utf8
5551     local lbkr = Babel.linebreaking.replacements[mode]
5552
5553     local word_head = head
5554
5555     while true do    %% for each subtext block
5556
5557         local w, wn, nw, lang = Babel.fetch_subtext[mode](word_head)
5558
5559         if Babel.debug then
5560             print()
5561             print('@@@@', w, nw)
5562         end
5563
5564         if nw == nil and w == '' then break end
5565
5566         if not lang then goto next end
5567         if not lbkr[lang] then goto next end
5568
5569         %% For each saved (pre|post)hyphenation. TODO. Reconsider how
5570         %% loops are nested.
5571         for k=1, #lbkr[lang] do
5572             local p = lbkr[lang][k].pattern
5573             local r = lbkr[lang][k].replace
5574
5575             if Babel.debug then
5576                 print('====', p, mode)
5577             end
5578
5579             %% This variable is set in some cases below to the first *byte*
5580             %% after the match, either as found by u.match (faster) or the
5581             %% computed position based on sc if w has changed.
5582             local last_match = 0
5583
5584             %% For every match.
5585             while true do
5586                 if Babel.debug then
5587                     print('-----')
5588                 end
5589                 local new    %% used when inserting and removing nodes
5590                 local refetch = false
5591
5592                 local matches = { u.match(w, p, last_match) }
5593                 if #matches < 2 then break end

```

```

5594
5595      %% Get and remove empty captures (with ())'s, which return a
5596      %% number with the position), and keep actual captures
5597      %% (from (...)), if any, in matches.
5598      local first = table.remove(matches, 1)
5599      local last = table.remove(matches, #matches)
5600      %% Non re-fetched substrings may contain \31, which separates
5601      %% subsubstrings.
5602      if string.find(w:sub(first, last-1), Babel.us_char) then break end
5603
5604      local save_last = last %% with A()BC()D, points to D
5605
5606      %% Fix offsets, from bytes to unicode. Explained above.
5607      first = u.len(w:sub(1, first-1)) + 1
5608      last = u.len(w:sub(1, last-1)) %% now last points to C
5609
5610      if Babel.debug then
5611          print(p)
5612          print('', 'sc', 'first', 'last', 'last_m', 'w')
5613      end
5614
5615      %% This loop traverses the matched substring and takes the
5616      %% corresponding action stored in the replacement list.
5617      %% sc = the position in substr nodes / string
5618      %% rc = the replacement table index
5619      local sc = first-1
5620      local rc = 0
5621      while rc < last-first+1 do %% for each replacement
5622          if Babel.debug then
5623              print('.....')
5624          end
5625          sc = sc + 1
5626          rc = rc + 1
5627          local crep = r[rc]
5628          local char_node = wn[sc]
5629          local char_base = char_node
5630          local end_replacement = false
5631
5632          if crep and crep.data then
5633              char_base = wn[crep.data+first-1]
5634          end
5635
5636          if Babel.debug then
5637              print('*', sc, first, last, last_match, w)
5638          end
5639
5640          if crep and next(crep) == nil then %% {}
5641              last_match = save_last
5642
5643          elseif crep == nil then %% remove
5644              node.remove(head, char_node)
5645              table.remove(wn, sc)
5646              w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
5647              last_match = utf8.offset(w, sc)
5648              sc = sc - 1 %% Nothing has been inserted
5649
5650          elseif mode == 1 and crep and (crep.pre or crep.no or crep.post) then
5651              local d = node.new(7, 0) %% (disc, discretionary)
5652              d.pre = Babel.str_to_nodes(crep.pre, matches, char_base)

```

```

5653         d.post = Babel.str_to_nodes(crep.post, matches, char_base)
5654         d.replace = Babel.str_to_nodes(crep.no, matches, char_base)
5655         d.attr = char_base.attr
5656         if crep.pre == nil then &% TeXbook p96
5657             d.penalty = crep.penalty or tex.hyphenpenalty
5658         else
5659             d.penalty = crep.penalty or tex.exhyphenpenalty
5660         end
5661         head, new = node.insert_before(head, char_node, d)
5662         end_replacement = true
5663
5664     elseif crep and crep.penalty then
5665         local d = node.new(14, 0) &% (penalty, userpenalty)
5666         d.attr = char_base.attr
5667         d.penalty = crep.penalty
5668         head, new = node.insert_before(head, char_node, d)
5669         end_replacement = true
5670
5671     elseif crep and crep.string then
5672         local str = crep.string(matches)
5673         if str == '' then &% Gather with nil
5674             refetch = true
5675             if sc == 1 then
5676                 word_head = char_node.next
5677             end
5678             head, new = node.remove(head, char_node)
5679         elseif char_node.id == 29 and u.len(str) == 1 then
5680             char_node.char = string.utfvalue(str)
5681             w = u.sub(w, 1, sc-1) .. str .. u.sub(w, sc+1)
5682             last_match = utf8.offset(w, sc+1)
5683         else
5684             refetch = true
5685             local n
5686             for s in string.utfvalues(str) do
5687                 if char_node.id == 7 then
5688                     &% TODO. Remove this limitation.
5689                     texio.write_nl('Automatic hyphens cannot be replaced, just removed.')
5690                 else
5691                     n = node.copy(char_base)
5692                 end
5693                 n.char = s
5694                 if sc == 1 then
5695                     head, new = node.insert_before(head, char_node, n)
5696                     word_head = new
5697                 else
5698                     node.insert_before(head, char_node, n)
5699                 end
5700             end
5701             node.remove(head, char_node)
5702         end &% string length
5703     end &% if char and char.string (ie replacement cases)
5704
5705     &% Shared by disc and penalty.
5706     if end_replacement then
5707         if sc == 1 then
5708             word_head = new
5709         end
5710         if crep.insert then
5711             last_match = save_last

```

```

5712         else
5713             node.remove(head, char_node)
5714             w = u.sub(w, 1, sc-1) .. Babel.us_char .. u.sub(w, sc+1)
5715             last_match = utf8.offset(w, sc)
5716         end
5717     end
5718 end    %% for each replacement
5719
5720 if Babel.debug then
5721     print('/', sc, first, last, last_match, w)
5722 end
5723
5724 %% TODO. refetch will be eventually unnecessary.
5725 if refetch then
5726     w, wn, nw, lang = Babel.fetch_subtext[mode](word_head)
5727 end
5728
5729     end    %% for match
5730 end    %% for patterns
5731
5732 ::next::
5733     word_head = nw
5734 end    %% for substring
5735 return head
5736 end
5737
5738 %% This table stores capture maps, numbered consecutively
5739 Babel.capture_maps = {}
5740
5741 %% The following functions belong to the next macro
5742 function Babel.capture_func(key, cap)
5743     local ret = "[" .. cap:gsub('{{[0-9]}}', "]]..m[%1]..[" .. "]"
5744     ret = ret:gsub('{{[0-9]}|([^\]]+)|(.-)}', Babel.capture_func_map)
5745     ret = ret:gsub("%[%[%]]%.", '')
5746     ret = ret:gsub("%.%[%[%]]%", '')
5747     return key .. [[=function(m) return ]] .. ret .. [[ end]]
5748 end
5749
5750 function Babel.capt_map(from, mapno)
5751     return Babel.capture_maps[mapno][from] or from
5752 end
5753
5754 %% Handle the {n|abc|ABC} syntax in captures
5755 function Babel.capture_func_map(capno, from, to)
5756     local froms = {}
5757     for s in string.utfcharacters(from) do
5758         table.insert(froms, s)
5759     end
5760     local cnt = 1
5761     table.insert(Babel.capture_maps, {})
5762     local mlen = table.getn(Babel.capture_maps)
5763     for s in string.utfcharacters(to) do
5764         Babel.capture_maps[mlen][froms[cnt]] = s
5765         cnt = cnt + 1
5766     end
5767     return "]]..Babel.capt_map(m[" .. capno .. "], " ..
5768         (mlen) .. ").." .. "["
5769 end
5770 }

```

Now the \TeX high level interface, which requires the function defined above for converting strings to functions returning a string. These functions handle the $\{n\}$ syntax. For example, $\text{pre}=\{1\}\{1\}$ - becomes `function(m) return m[1]..m[1]..'-' end`, where m are the matches returned after applying the pattern. With a mapped capture the functions are similar to `function(m) return Babel.capt_map(m[1],1) end`, where the last argument identifies the mapping to be applied to $m[1]$. The way it is carried out is somewhat tricky, but the effect is not dissimilar to `lua load – save the code as string in a \TeX macro, and expand this macro at the appropriate place. As \directlua does not take into account the current catcode of @, we just avoid this character in macro names (which explains the internal group, too).`

```

5771 \catcode`\#=6
5772 \gdef\babelposthyphenation#1#2#3{&%
5773   \bbl@activateposthyphen
5774   \begingroup
5775     \def\babeltempa{\bbl@add@list\babeltempb}&%
5776     \let\babeltempb\@empty
5777     \bbl@foreach{#3}{&%
5778       \bbl@ifsamestring{##1}{remove}&%
5779       {\bbl@add@list\babeltempb{nil}}&%
5780       {\directlua{
5781         local rep = [[#1]]
5782         rep = rep:gsub('^%s*(insert)%s*', 'insert = true, ')
5783         rep = rep:gsub(' (no)%s*=%s*([^\s,]*)', Babel.capture_func)
5784         rep = rep:gsub(' (pre)%s*=%s*([^\s,]*)', Babel.capture_func)
5785         rep = rep:gsub(' (post)%s*=%s*([^\s,]*)', Babel.capture_func)
5786         rep = rep:gsub('(string)%s*=%s*([^\s,]*)', Babel.capture_func)
5787         tex.print([[\\string\babeltempa{}} .. rep .. [[{}]])
5788       }}&%
5789     \directlua{
5790       local lbkr = Babel.linebreaking.replacements[1]
5791       local u = unicode.utf8
5792       &% Convert pattern:
5793       local patt = string.gsub([=[#2]=], '%s', '')
5794       if not u.find(patt, '()', nil, true) then
5795         patt = '()' .. patt .. '()'
5796       end
5797       patt = string.gsub(patt, '%(%)^', '^()')
5798       patt = string.gsub(patt, '%$(%)', '()$')
5799       patt = u.gsub(patt, '{(.)}',
5800         function (n)
5801           return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5802         end)
5803       lbkr[\\the\\csname l@#1\\endcsname] = lbkr[\\the\\csname l@#1\\endcsname] or {}
5804       table.insert(lbkr[\\the\\csname l@#1\\endcsname],
5805         { pattern = patt, replace = { \babeltempb } })
5806     }&%
5807   \endgroup}
5808 % TODO. Copypaste pattern.
5809 \gdef\babelprehyphenation#1#2#3{&%
5810   \bbl@activateprehyphen
5811   \begingroup
5812     \def\babeltempa{\bbl@add@list\babeltempb}&%
5813     \let\babeltempb\@empty
5814     \bbl@foreach{#3}{&%
5815       \bbl@ifsamestring{##1}{remove}&%
5816       {\bbl@add@list\babeltempb{nil}}&%
5817       {\directlua{
5818         local rep = [[#1]]

```



```

5819         rep = rep:gsub('^%s*(insert)%s*', ' ', 'insert = true, ')
5820         rep = rep:gsub('(string)%s*=%s*([^\s,]*)', Babel.capture_func)
5821         tex.print([[\\string\\babeltempa{}}] .. rep .. [[]}}]])
5822     }}&%
5823     \\directlua{
5824         local lbkr = Babel.linebreaking.replacements[0]
5825         local u = unicode.utf8
5826         &% Convert pattern:
5827         local patt = string.gsub([==[#2]==], '%s', ' ')
5828         if not u.find(patt, '()', nil, true) then
5829             patt = '()' .. patt .. '()'
5830         end
5831         &% patt = string.gsub(patt, '%(%)^', '^()')
5832         &% patt = string.gsub(patt, '([^\s])%$%$', '%1()$')
5833         patt = u.gsub(patt, '{(.)}',
5834             function (n)
5835                 return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5836             end)
5837         lbkr[\\the\\csname bbl@id@@#1\\endcsname] = lbkr[\\the\\csname bbl@id@@#1\\endcsname] or {}
5838         table.insert(lbkr[\\the\\csname bbl@id@@#1\\endcsname],
5839             { pattern = patt, replace = { \\babeltempb } })
5840     }&%
5841     \\endgroup}
5842 \\endgroup
5843 \\def\\bbl@activateposthyphen{%
5844     \\let\\bbl@activateposthyphen\\relax
5845     \\directlua{
5846         Babel.linebreaking.add_after(Babel.post_hyphenate_replace)
5847     }}
5848 \\def\\bbl@activateprehyphen{%
5849     \\let\\bbl@activateprehyphen\\relax
5850     \\directlua{
5851         Babel.linebreaking.add_before(Babel.pre_hyphenate_replace)
5852     }}

```

13.7 Layout

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

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

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

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

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

```

5853 \\bbl@trace{Redefinitions for bidi layout}
5854 \\ifx\\@eqnnum\\undefined\\else
5855     \\ifx\\bbl@attr@dir\\undefined\\else
5856         \\edef\\@eqnnum{%
5857             \\unexpanded{\\ifcase\\bbl@attr@dir\\else\\bbl@textdir\\@ne\\fi}%
5858             \\unexpanded\\expandafter{\\@eqnnum}}%
5859     \\fi
5860 \\fi

```

```

5861 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
5862 \ifnum\bbl@bidimode>\z@
5863 \def\bbl@nextfake#1{% non-local changes, use always inside a group!
5864 \bbl@exp{%
5865 \mathdir\the\bodydir
5866 #1% Once entered in math, set boxes to restore values
5867 \<ifmode>%
5868 \everyvbox{%
5869 \the\everyvbox
5870 \bodydir\the\bodydir
5871 \mathdir\the\mathdir
5872 \everyhbox{\the\everyhbox}%
5873 \everyvbox{\the\everyvbox}}%
5874 \everyhbox{%
5875 \the\everyhbox
5876 \bodydir\the\bodydir
5877 \mathdir\the\mathdir
5878 \everyhbox{\the\everyhbox}%
5879 \everyvbox{\the\everyvbox}}%
5880 \<fi>}}%
5881 \def\@hangfrom#1{%
5882 \setbox\@tempboxa\hbox{{#1}}%
5883 \hangindent\wd\@tempboxa
5884 \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
5885 \shapemode\@ne
5886 \fi
5887 \noindent\box\@tempboxa}
5888 \fi
5889 \IfBabelLayout{tabular}
5890 {\let\bbl@OL@tabular\@tabular
5891 \bbl@replace\@tabular{$}\{\bbl@nextfake$}%
5892 \let\bbl@NL@tabular\@tabular
5893 \AtBeginDocument{%
5894 \ifx\bbl@NL@tabular\@tabular\else
5895 \bbl@replace\@tabular{$}\{\bbl@nextfake$}%
5896 \let\bbl@NL@tabular\@tabular
5897 \fi}}
5898 {}
5899 \IfBabelLayout{lists}
5900 {\let\bbl@OL@list\list
5901 \bbl@sreplace\list{\parshape}\{\bbl@listparshape}%
5902 \let\bbl@NL@list\list
5903 \def\bbl@listparshape#1#2#3{%
5904 \parshape #1 #2 #3 %
5905 \ifnum\bbl@getluadir{page}=\bbl@getluadir{par}\else
5906 \shapemode\tw@
5907 \fi}}
5908 {}
5909 \IfBabelLayout{graphics}
5910 {\let\bbl@pictresetdir\relax
5911 \def\bbl@pictsetdir{%
5912 \ifcase\bbl@thetextdir
5913 \let\bbl@pictresetdir\relax
5914 \else
5915 \textdir TLT\relax
5916 \def\bbl@pictresetdir{\textdir TRT\relax}%
5917 \fi}%
5918 \let\bbl@OL@picture\@picture
5919 \let\bbl@OL@put\put

```

```

5920 \bbl@sreplace\@picture{\hskip-}\{\bbl@pictsetdir\hskip-}%
5921 \def\put(#1,#2)#3{% Not easy to patch. Better redefine.
5922   \@killglue
5923   \raise#2\unitlength
5924   \hb@xt@\z@{\kern#1\unitlength{\bbl@pictresetdir#3}\hss}}%
5925 \AtBeginDocument
5926   {\ifx\tikz@atbegin@node\undefined\else
5927     \let\bbl@OL@pgfpicture\pgfpicture
5928     \bbl@sreplace\pgfpicture{\pgfpicturetrue}%
5929     {\bbl@pictsetdir\pgfpicturetrue}%
5930     \bbl@add\pgfsys@beginpicture{\bbl@pictsetdir}%
5931     \bbl@add\tikz@atbegin@node{\bbl@pictresetdir}%
5932     \fi}}
5933 {}

```

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

```

5934 \IfBabelLayout{counters}%
5935   {\let\bbl@OL@@textsuperscript\textsuperscript
5936     \bbl@sreplace\textsuperscript{\m@th}\{\m@th\mathdir\pagedir}%
5937     \let\bbl@latinarabic=\@arabic
5938     \let\bbl@OL@@arabic\@arabic
5939     \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
5940     \@ifpackagewith{babel}{bidi=default}%
5941     {\let\bbl@asciroman=\@roman
5942       \let\bbl@OL@@roman\@roman
5943       \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
5944       \let\bbl@asciiRoman=\@Roman
5945       \let\bbl@OL@@roman\@Roman
5946       \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
5947       \let\bbl@OL@labelenumii\labelenumii
5948       \def\labelenumii{}\theenumii}%
5949       \let\bbl@OL@p@enumiii\p@enumiii
5950       \def\p@enumiii{\p@enumii}\theenumii{}\{\}\{\}}
5951   }{Footnote changes}
5952 \IfBabelLayout{footnotes}%
5953   {\let\bbl@OL@footnote\footnote
5954     \BabelFootnote\footnote\language\{\}\}%
5955     \BabelFootnote\localfootnote\language\{\}\}%
5956     \BabelFootnote\mainfootnote\{\}\{\}\{\}}
5957   {}

```

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

```

5958 \IfBabelLayout{extras}%
5959   {\let\bbl@OL@underline\underline
5960     \bbl@sreplace\underline{\$@@@underline}\{\bbl@nextfake\$@@@underline}%
5961     \let\bbl@OL@LaTeX2e\LaTeX2e
5962     \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
5963       \if b\expandafter\@car\@series\@nil\boldmath\fi
5964       \babelsublr{%
5965         \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}%
5966   {}
5967 }/luatex}

```

13.8 Auto bidi with basic and basic-r

The file `babel-data-bidi.lua` currently only contains data. It is a large and boring file and it is not shown here (see the generated file), but here is a sample:

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

For the meaning of these codes, see the Unicode standard.

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

Arrrgh!! The UAX#9 algorithm is too deeply entrenched in the assumption of batch-style processing [...]. May the fleas of a thousand camels infest the armpits of those who design supposedly general-purpose algorithms by looking at their own implementations, and fail to consider other possible implementations!

Well, it took me some time to guess what the batch rules in UAX#9 actually mean (in other word, *what* they do and *why*, and not only *how*), but I think (or I hope) I've managed to understand them.

In some sense, there are two bidi modes, one for numbers, and the other for text. Furthermore, setting just the direction in R text is not enough, because there are actually *two* R modes (set explicitly in Unicode with RLM and ALM). In babel the `dir` is set by a higher protocol based on the language/script, which in turn sets the correct `dir` (`<l>`, `<r>` or `<al>`).

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

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

```
5968 (*basic-r)
5969 Babel = Babel or {}
5970
5971 Babel.bidi_enabled = true
5972
5973 require('babel-data-bidi.lua')
5974
5975 local characters = Babel.characters
5976 local ranges = Babel.ranges
5977
5978 local DIR = node.id("dir")
5979
5980 local function dir_mark(head, from, to, outer)
5981   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
```

```

5982 local d = node.new(DIR)
5983 d.dir = '+' .. dir
5984 node.insert_before(head, from, d)
5985 d = node.new(DIR)
5986 d.dir = '-' .. dir
5987 node.insert_after(head, to, d)
5988 end
5989
5990 function Babel.bidi(head, ispar)
5991   local first_n, last_n      -- first and last char with nums
5992   local last_es              -- an auxiliary 'last' used with nums
5993   local first_d, last_d      -- first and last char in L/R block
5994   local dir, dir_real

```

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

```

5995 local strong = ('TRT' == tex.pardir) and 'r' or 'l'
5996 local strong_lr = (strong == 'l') and 'l' or 'r'
5997 local outer = strong
5998
5999 local new_dir = false
6000 local first_dir = false
6001 local inmath = false
6002
6003 local last_lr
6004
6005 local type_n = ''
6006
6007 for item in node.traverse(head) do
6008
6009   -- three cases: glyph, dir, otherwise
6010   if item.id == node.id'glyph'
6011     or (item.id == 7 and item.subtype == 2) then
6012
6013     local itemchar
6014     if item.id == 7 and item.subtype == 2 then
6015       itemchar = item.replace.char
6016     else
6017       itemchar = item.char
6018     end
6019     local chardata = characters[itemchar]
6020     dir = chardata and chardata.d or nil
6021     if not dir then
6022       for nn, et in ipairs(ranges) do
6023         if itemchar < et[1] then
6024           break
6025         elseif itemchar <= et[2] then
6026           dir = et[3]
6027           break
6028         end
6029       end
6030     end
6031     dir = dir or 'l'
6032     if inmath then dir = ('TRT' == tex.mathdir) and 'r' or 'l' end

```

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

then. This is not exactly true, as the math mode may insert explicit dirs in the node list, so, for the moment there is a hack by brute force (just above).

```

6033     if new_dir then
6034         attr_dir = 0
6035         for at in node.traverse(item.attr) do
6036             if at.number == luatexbase.registernumber'bbl@attr@dir' then
6037                 attr_dir = at.value % 3
6038             end
6039         end
6040         if attr_dir == 1 then
6041             strong = 'r'
6042         elseif attr_dir == 2 then
6043             strong = 'al'
6044         else
6045             strong = 'l'
6046         end
6047         strong_lr = (strong == 'l') and 'l' or 'r'
6048         outer = strong_lr
6049         new_dir = false
6050     end
6051
6052     if dir == 'nsm' then dir = strong end          -- W1

```

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

```

6053     dir_real = dir          -- We need dir_real to set strong below
6054     if dir == 'al' then dir = 'r' end -- W3

```

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

```

6055     if strong == 'al' then
6056         if dir == 'en' then dir = 'an' end          -- W2
6057         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
6058         strong_lr = 'r'                             -- W3
6059     end

```

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

```

6060     elseif item.id == node.id'dir' and not inmath then
6061         new_dir = true
6062         dir = nil
6063     elseif item.id == node.id'math' then
6064         inmath = (item.subtype == 0)
6065     else
6066         dir = nil          -- Not a char
6067     end

```

Numbers in R mode. A sequence of `<en>`, `<et>`, `<an>`, `<es>` and `<cs>` is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, ie, a non-char), the `texdir` is set. This means you cannot insert, say, a `whatsit`, but this is what I would expect (with `luacolor` you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only `<an>` is relevant if `<al>`.

```

6068     if dir == 'en' or dir == 'an' or dir == 'et' then
6069         if dir ~= 'et' then
6070             type_n = dir
6071         end
6072         first_n = first_n or item
6073         last_n = last_es or item

```

```

6074     last_es = nil
6075   elseif dir == 'es' and last_n then -- W3+W6
6076     last_es = item
6077   elseif dir == 'cs' then           -- it's right - do nothing
6078   elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
6079     if strong_lr == 'r' and type_n ~= '' then
6080       dir_mark(head, first_n, last_n, 'r')
6081     elseif strong_lr == 'l' and first_d and type_n == 'an' then
6082       dir_mark(head, first_n, last_n, 'r')
6083       dir_mark(head, first_d, last_d, outer)
6084       first_d, last_d = nil, nil
6085     elseif strong_lr == 'l' and type_n ~= '' then
6086       last_d = last_n
6087     end
6088     type_n = ''
6089     first_n, last_n = nil, nil
6090   end

```

R text in L, or L text in R. Order of dir_ mark's are relevant: d goes outside n, and therefore it's emitted after. See dir_mark to understand why (but is the nesting actually necessary or is a flat dir structure enough?). Only L, R (and AL) chars are taken into account – everything else, including spaces, whatsits, etc., are ignored:

```

6091   if dir == 'l' or dir == 'r' then
6092     if dir ~= outer then
6093       first_d = first_d or item
6094       last_d = item
6095     elseif first_d and dir ~= strong_lr then
6096       dir_mark(head, first_d, last_d, outer)
6097       first_d, last_d = nil, nil
6098     end
6099   end

```

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

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

```

6100   if dir and not last_lr and dir ~= 'l' and outer == 'r' then
6101     item.char = characters[item.char] and
6102               characters[item.char].m or item.char
6103   elseif (dir or new_dir) and last_lr ~= item then
6104     local mir = outer .. strong_lr .. (dir or outer)
6105     if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
6106       for ch in node.traverse(node.next(last_lr)) do
6107         if ch == item then break end
6108         if ch.id == node.id'glyph' and characters[ch.char] then
6109           ch.char = characters[ch.char].m or ch.char
6110         end
6111       end
6112     end
6113   end

```

Save some values for the next iteration. If the current node is 'dir', open a new sequence. Since dir could be changed, strong is set with its real value (dir_real).

```

6114   if dir == 'l' or dir == 'r' then
6115     last_lr = item
6116     strong = dir_real           -- Don't search back - best save now
6117     strong_lr = (strong == 'l') and 'l' or 'r'

```

```

6118     elseif new_dir then
6119         last_lr = nil
6120     end
6121 end

```

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

```

6122 if last_lr and outer == 'r' then
6123     for ch in node.traverse_id(node.id('glyph', node.next(last_lr))) do
6124         if characters[ch.char] then
6125             ch.char = characters[ch.char].m or ch.char
6126         end
6127     end
6128 end
6129 if first_n then
6130     dir_mark(head, first_n, last_n, outer)
6131 end
6132 if first_d then
6133     dir_mark(head, first_d, last_d, outer)
6134 end

```

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

```

6135 return node.prev(head) or head
6136 end
6137 </basic-r>

```

And here the Lua code for bidi=basic:

```

6138 (*basic)
6139 Babel = Babel or {}
6140
6141 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
6142
6143 Babel.fontmap = Babel.fontmap or {}
6144 Babel.fontmap[0] = {}      -- l
6145 Babel.fontmap[1] = {}      -- r
6146 Babel.fontmap[2] = {}      -- al/an
6147
6148 Babel.bidi_enabled = true
6149 Babel.mirroring_enabled = true
6150
6151 require('babel-data-bidi.lua')
6152
6153 local characters = Babel.characters
6154 local ranges = Babel.ranges
6155
6156 local DIR = node.id('dir')
6157 local GLYPH = node.id('glyph')
6158
6159 local function insert_implicit(head, state, outer)
6160     local new_state = state
6161     if state.sim and state.eim and state.sim ~= state.eim then
6162         dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
6163         local d = node.new(DIR)
6164         d.dir = '+' .. dir
6165         node.insert_before(head, state.sim, d)
6166         local d = node.new(DIR)
6167         d.dir = '-' .. dir
6168         node.insert_after(head, state.eim, d)
6169     end

```



```

6170 new_state.sim, new_state.eim = nil, nil
6171 return head, new_state
6172 end
6173
6174 local function insert_numeric(head, state)
6175   local new
6176   local new_state = state
6177   if state.san and state.ean and state.san ~= state.ean then
6178     local d = node.new(DIR)
6179     d.dir = '+TLT'
6180     _, new = node.insert_before(head, state.san, d)
6181     if state.san == state.sim then state.sim = new end
6182     local d = node.new(DIR)
6183     d.dir = '-TLT'
6184     _, new = node.insert_after(head, state.ean, d)
6185     if state.ean == state.eim then state.eim = new end
6186   end
6187   new_state.san, new_state.ean = nil, nil
6188   return head, new_state
6189 end
6190
6191 -- TODO - \hbox with an explicit dir can lead to wrong results
6192 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
6193 -- was made to improve the situation, but the problem is the 3-dir
6194 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
6195 -- well.
6196
6197 function Babel.bidi(head, ispar, hdir)
6198   local d -- d is used mainly for computations in a loop
6199   local prev_d = ''
6200   local new_d = false
6201
6202   local nodes = {}
6203   local outer_first = nil
6204   local inmath = false
6205
6206   local glue_d = nil
6207   local glue_i = nil
6208
6209   local has_en = false
6210   local first_et = nil
6211
6212   local ATDIR = luatexbase.registernumber'bbl@attr@dir'
6213
6214   local save_outer
6215   local temp = node.get_attribute(head, ATDIR)
6216   if temp then
6217     temp = temp % 3
6218     save_outer = (temp == 0 and 'l') or
6219                 (temp == 1 and 'r') or
6220                 (temp == 2 and 'al')
6221   elseif ispar then -- Or error? Shouldn't happen
6222     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
6223   else -- Or error? Shouldn't happen
6224     save_outer = ('TRT' == hdir) and 'r' or 'l'
6225   end
6226   -- when the callback is called, we are just _after_ the box,
6227   -- and the textdir is that of the surrounding text
6228   -- if not ispar and hdir ~= tex.textdir then

```

```

6229 -- save_outer = ('TRT' == hdir) and 'r' or 'l'
6230 -- end
6231 local outer = save_outer
6232 local last = outer
6233 -- 'al' is only taken into account in the first, current loop
6234 if save_outer == 'al' then save_outer = 'r' end
6235
6236 local fontmap = Babel.fontmap
6237
6238 for item in node.traverse(head) do
6239
6240     -- In what follows, #node is the last (previous) node, because the
6241     -- current one is not added until we start processing the neutrals.
6242
6243     -- three cases: glyph, dir, otherwise
6244     if item.id == GLYPH
6245         or (item.id == 7 and item.subtype == 2) then
6246
6247         local d_font = nil
6248         local item_r
6249         if item.id == 7 and item.subtype == 2 then
6250             item_r = item.replace -- automatic discs have just 1 glyph
6251         else
6252             item_r = item
6253         end
6254         local chardata = characters[item_r.char]
6255         d = chardata and chardata.d or nil
6256         if not d or d == 'nsm' then
6257             for nn, et in ipairs(ranges) do
6258                 if item_r.char < et[1] then
6259                     break
6260                 elseif item_r.char <= et[2] then
6261                     if not d then d = et[3]
6262                     elseif d == 'nsm' then d_font = et[3]
6263                     end
6264                     break
6265                 end
6266             end
6267         end
6268         d = d or 'l'
6269
6270         -- A short 'pause' in bidi for mapfont
6271         d_font = d_font or d
6272         d_font = (d_font == 'l' and 0) or
6273             (d_font == 'nsm' and 0) or
6274             (d_font == 'r' and 1) or
6275             (d_font == 'al' and 2) or
6276             (d_font == 'an' and 2) or nil
6277         if d_font and fontmap and fontmap[d_font][item_r.font] then
6278             item_r.font = fontmap[d_font][item_r.font]
6279         end
6280
6281         if new_d then
6282             table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6283             if inmath then
6284                 attr_d = 0
6285             else
6286                 attr_d = node.get_attribute(item, ATDIR)
6287                 attr_d = attr_d % 3

```

```

6288         end
6289         if attr_d == 1 then
6290             outer_first = 'r'
6291             last = 'r'
6292         elseif attr_d == 2 then
6293             outer_first = 'r'
6294             last = 'al'
6295         else
6296             outer_first = 'l'
6297             last = 'l'
6298         end
6299         outer = last
6300         has_en = false
6301         first_et = nil
6302         new_d = false
6303     end
6304
6305     if glue_d then
6306         if (d == 'l' and 'l' or 'r') ~= glue_d then
6307             table.insert(nodes, {glue_i, 'on', nil})
6308         end
6309         glue_d = nil
6310         glue_i = nil
6311     end
6312
6313     elseif item.id == DIR then
6314         d = nil
6315         new_d = true
6316
6317     elseif item.id == node.id'glue' and item.subtype == 13 then
6318         glue_d = d
6319         glue_i = item
6320         d = nil
6321
6322     elseif item.id == node.id'math' then
6323         inmath = (item.subtype == 0)
6324
6325     else
6326         d = nil
6327     end
6328
6329     -- AL <= EN/ET/ES      -- W2 + W3 + W6
6330     if last == 'al' and d == 'en' then
6331         d = 'an'          -- W3
6332     elseif last == 'al' and (d == 'et' or d == 'es') then
6333         d = 'on'          -- W6
6334     end
6335
6336     -- EN + CS/ES + EN      -- W4
6337     if d == 'en' and #nodes >= 2 then
6338         if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
6339             and nodes[#nodes-1][2] == 'en' then
6340             nodes[#nodes][2] = 'en'
6341         end
6342     end
6343
6344     -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
6345     if d == 'an' and #nodes >= 2 then
6346         if (nodes[#nodes][2] == 'cs')

```

```

6347         and nodes[#nodes-1][2] == 'an' then
6348             nodes[#nodes][2] = 'an'
6349         end
6350     end
6351
6352     -- ET/EN             -- W5 + W7->l / W6->on
6353     if d == 'et' then
6354         first_et = first_et or (#nodes + 1)
6355     elseif d == 'en' then
6356         has_en = true
6357         first_et = first_et or (#nodes + 1)
6358     elseif first_et then      -- d may be nil here !
6359         if has_en then
6360             if last == 'l' then
6361                 temp = 'l'      -- W7
6362             else
6363                 temp = 'en'     -- W5
6364             end
6365         else
6366             temp = 'on'        -- W6
6367         end
6368         for e = first_et, #nodes do
6369             if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6370         end
6371         first_et = nil
6372         has_en = false
6373     end
6374
6375     if d then
6376         if d == 'al' then
6377             d = 'r'
6378             last = 'al'
6379         elseif d == 'l' or d == 'r' then
6380             last = d
6381         end
6382         prev_d = d
6383         table.insert(nodes, {item, d, outer_first})
6384     end
6385
6386     outer_first = nil
6387
6388 end
6389
6390 -- TODO -- repeated here in case EN/ET is the last node. Find a
6391 -- better way of doing things:
6392 if first_et then      -- dir may be nil here !
6393     if has_en then
6394         if last == 'l' then
6395             temp = 'l'      -- W7
6396         else
6397             temp = 'en'     -- W5
6398         end
6399     else
6400         temp = 'on'        -- W6
6401     end
6402     for e = first_et, #nodes do
6403         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6404     end
6405 end

```

```

6406
6407 -- dummy node, to close things
6408 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6409
6410 ----- NEUTRAL -----
6411
6412 outer = save_outer
6413 last = outer
6414
6415 local first_on = nil
6416
6417 for q = 1, #nodes do
6418     local item
6419
6420     local outer_first = nodes[q][3]
6421     outer = outer_first or outer
6422     last = outer_first or last
6423
6424     local d = nodes[q][2]
6425     if d == 'an' or d == 'en' then d = 'r' end
6426     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
6427
6428     if d == 'on' then
6429         first_on = first_on or q
6430     elseif first_on then
6431         if last == d then
6432             temp = d
6433         else
6434             temp = outer
6435         end
6436         for r = first_on, q - 1 do
6437             nodes[r][2] = temp
6438             item = nodes[r][1] -- MIRRORING
6439             if Babel.mirroring_enabled and item.id == GLYPH
6440                 and temp == 'r' and characters[item.char] then
6441                 local font_mode = font.fonts[item.font].properties.mode
6442                 if font_mode ~= 'harf' and font_mode ~= 'plug' then
6443                     item.char = characters[item.char].m or item.char
6444                 end
6445             end
6446         end
6447         first_on = nil
6448     end
6449
6450     if d == 'r' or d == 'l' then last = d end
6451 end
6452
6453 ----- IMPLICIT, REORDER -----
6454
6455 outer = save_outer
6456 last = outer
6457
6458 local state = {}
6459 state.has_r = false
6460
6461 for q = 1, #nodes do
6462
6463     local item = nodes[q][1]
6464

```

```

6465     outer = nodes[q][3] or outer
6466
6467     local d = nodes[q][2]
6468
6469     if d == 'nsm' then d = last end           -- W1
6470     if d == 'en' then d = 'an' end
6471     local isdir = (d == 'r' or d == 'l')
6472
6473     if outer == 'l' and d == 'an' then
6474         state.san = state.san or item
6475         state.ean = item
6476     elseif state.san then
6477         head, state = insert_numeric(head, state)
6478     end
6479
6480     if outer == 'l' then
6481         if d == 'an' or d == 'r' then        -- im -> implicit
6482             if d == 'r' then state.has_r = true end
6483             state.sim = state.sim or item
6484             state.eim = item
6485         elseif d == 'l' and state.sim and state.has_r then
6486             head, state = insert_implicit(head, state, outer)
6487         elseif d == 'l' then
6488             state.sim, state.eim, state.has_r = nil, nil, false
6489         end
6490     else
6491         if d == 'an' or d == 'l' then
6492             if nodes[q][3] then -- nil except after an explicit dir
6493                 state.sim = item -- so we move sim 'inside' the group
6494             else
6495                 state.sim = state.sim or item
6496             end
6497             state.eim = item
6498         elseif d == 'r' and state.sim then
6499             head, state = insert_implicit(head, state, outer)
6500         elseif d == 'r' then
6501             state.sim, state.eim = nil, nil
6502         end
6503     end
6504
6505     if isdir then
6506         last = d           -- Don't search back - best save now
6507     elseif d == 'on' and state.san then
6508         state.san = state.san or item
6509         state.ean = item
6510     end
6511
6512 end
6513
6514 return node.prev(head) or head
6515 end
6516 </basic>

```

14 Data for CJK

It is a boring file and it is not shown here (see the generated file), but here is a sample:

```
[0x0021]={c='ex'},
[0x0024]={c='pr'},
[0x0025]={c='po'},
[0x0028]={c='op'},
[0x0029]={c='cp'},
[0x002B]={c='pr'},
```

For the meaning of these codes, see the Unicode standard.

15 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.

```
6517 < *nil>
6518 \ProvidesLanguage{nil}[\<<date>> \<<version>> Nil language]
6519 \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.

```
6520 \ifx\l@nil\undefined
6521 \newlanguage\l@nil
6522 \@namedef{bbl@hyphendata@the\l@nil}{\{}}% Remove warning
6523 \let\bbl@elt\relax
6524 \edef\bbl@languages{% Add it to the list of languages
6525 \bbl@languages\bbl@elt{nil}{\the\l@nil}{\{}}
6526 \fi
```

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

```
6527 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}
```

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

```
\captionnil
\datenil
6528 \let\captionnil\@empty
6529 \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.

```
6530 \ldf@finish{nil}
6531 < /nil>
```

16 Support for Plain T_EX (plain.def)

16.1 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:

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).

```
6532 (*bplain | blplain)
6533 \catcode`\{=1 % left brace is begin-group character
6534 \catcode`\}=2 % right brace is end-group character
6535 \catcode`\#=6 % hash mark is macro parameter character
```

```
6536 \openin 0 hyphen.cfg
6537 \ifeof0
6538 \else
6539   \let\input
```

```

6540 \def\input #1 {%
6541   \let\input\@
6542   \a hyphen.cfg
6543   \let\@undefined
6544 }
6545 \fi
6546 </bplain | bplain>

```

```
6547 <bplain>\a plain.tex
6548 <blplain>\a lplain.tex
```

```
6549 <bplain>\def\fmtname{babel-plain}
6550 <blplain>\def\fmtname{babel-lplain}
```

16.2 Emulating some L^AT_EX features

```

6551 \langle\langle *Emulate LaTeX \rangle\rangle \equiv
6552 % == Code for plain ==
6553 \def\@empty{}
6554 \def\loadlocalcfg#1{%
6555   \openin0#1.cfg

```



```

6556 \ifeof0
6557 \closein0
6558 \else
6559 \closein0
6560 {\immediate\write16{*****}%
6561 \immediate\write16{* Local config file #1.cfg used}%
6562 \immediate\write16{*}%
6563 }
6564 \input #1.cfg\relax
6565 \fi
6566 \@endofldf}

```

16.3 General tools

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

```

6567 \long\def\@firstofone#1{#1}
6568 \long\def\@firstoftwo#1#2{#1}
6569 \long\def\@secondoftwo#1#2{#2}
6570 \def\@nnil{\@nil}
6571 \def\@gobbletwo#1#2{}
6572 \def\@ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
6573 \def\@star@or@long#1{%
6574 \@ifstar
6575 {\let\l@ngrel@x\relax#1}%
6576 {\let\l@ngrel@x\long#1}}
6577 \let\l@ngrel@x\relax
6578 \def\@car#1#2\@nil{#1}
6579 \def\@cdr#1#2\@nil{#2}
6580 \let\@typeset@protect\relax
6581 \let\protected@edef\edef
6582 \long\def\@gobble#1{}
6583 \edef\@backslashchar{\expandafter\@gobble\string\}
6584 \def\strip@prefix#1>{}
6585 \def\g@addto@macro#1#2{%
6586 \toks@\expandafter{#1#2}%
6587 \xdef#1{\the\toks@}}
6588 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
6589 \def\@nameuse#1{\csname #1\endcsname}
6590 \def\@ifundefined#1{%
6591 \expandafter\ifx\csname#1\endcsname\relax
6592 \expandafter\@firstoftwo
6593 \else
6594 \expandafter\@secondoftwo
6595 \fi}
6596 \def\@expandtwoargs#1#2#3{%
6597 \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
6598 \def\zap@space#1 #2{%
6599 #1%
6600 \ifx#2\@empty\else\expandafter\zap@space\fi
6601 #2}
6602 \let\bbl@trace\@gobble

```

$\LaTeX 2_{\epsilon}$ has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

6603 \ifx\@preamblecmds\undefined
6604 \def\@preamblecmds{}
6605 \fi
6606 \def\@onlypreamble#1{%

```

```

6607 \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
6608 \@preamblecmds\do#1}}
6609 \@onlypreamble\@onlypreamble

```

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

```

6610 \def\begindocument{%
6611 \@begindocumenthook
6612 \global\let\@begindocumenthook\@undefined
6613 \def\do##1{\global\let##1\@undefined}%
6614 \@preamblecmds
6615 \global\let\do\noexpand}

6616 \ifx\@begindocumenthook\@undefined
6617 \def\@begindocumenthook{}
6618 \fi
6619 \@onlypreamble\@begindocumenthook
6620 \def\AtBeginDocument{\g@addto@macro\@begindocumenthook}

```

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

```

6621 \def\AtEndOfPackage#1{\g@addto@macro\@endoflfd{#1}}
6622 \@onlypreamble\AtEndOfPackage
6623 \def\@endoflfd{}
6624 \@onlypreamble\@endoflfd
6625 \let\bbl@afterlang\@empty
6626 \chardef\bbl@opt@hyphenmap\z@

```

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

```

6627 \catcode`\&=\z@
6628 \ifx&\if@files\@undefined
6629 \expandafter\let\csname if@files\expandafter\endcsname
6630 \csname iffalse\endcsname
6631 \fi
6632 \catcode`\&=4

```

Mimick \LaTeX 's commands to define control sequences.

```

6633 \def\newcommand{\@star@or@long\new@command}
6634 \def\new@command#1{%
6635 \@testopt{\@newcommand#1}0}
6636 \def\@newcommand#1[#2]{%
6637 \@ifnextchar [{\@xargdef#1[#2]}%
6638 {\@argdef#1[#2]}}
6639 \long\def\@argdef#1[#2]#3{%
6640 \@yargdef#1\@ne{#2}{#3}}
6641 \long\def\@xargdef#1[#2][#3]#4{%
6642 \expandafter\def\expandafter#1\expandafter{%
6643 \expandafter\@protected@testopt\expandafter #1%
6644 \csname\string#1\expandafter\endcsname{#3}}%
6645 \expandafter\@yargdef \csname\string#1\endcsname
6646 \tw@{#2}{#4}}
6647 \long\def\@yargdef#1#2#3{%
6648 \@tempcnta#3\relax
6649 \advance \@tempcnta \@ne
6650 \let\@hash@\relax
6651 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
6652 \@tempcntb #2%

```

```

6653 \@whilenum\@tempcntb <\@tempcnta
6654 \do{%
6655 \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
6656 \advance\@tempcntb \@ne}%
6657 \let\@hash@##%
6658 \l@ngrel@x\expandafter\def\expandafter#1\reserved@a}
6659 \def\providecommand{\@star@or@long\provide@command}
6660 \def\provide@command#1{%
6661 \begingroup
6662 \escapechar\m@ne\xdef\@gtempa{\string#1}}%
6663 \endgroup
6664 \expandafter\ifundefined\@gtempa
6665 {\def\reserved@a{\new@command#1}}%
6666 {\let\reserved@a\relax
6667 \def\reserved@a{\new@command\reserved@a}}%
6668 \reserved@a}%
6669 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
6670 \def\declare@robustcommand#1{%
6671 \edef\reserved@a{\string#1}%
6672 \def\reserved@b{#1}%
6673 \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
6674 \edef#1{%
6675 \ifx\reserved@a\reserved@b
6676 \noexpand\x@protect
6677 \noexpand#1%
6678 \fi
6679 \noexpand\protect
6680 \expandafter\noexpand\csname
6681 \expandafter\@gobble\string#1 \endcsname
6682 }%
6683 \expandafter\new@command\csname
6684 \expandafter\@gobble\string#1 \endcsname
6685 }
6686 \def\x@protect#1{%
6687 \ifx\protect\@typeset@protect\else
6688 \@x@protect#1%
6689 \fi
6690 }
6691 \catcode`\&=\z@ % Trick to hide conditionals
6692 \def\@x@protect#1&fi#2#3{&fi\protect#1}

```

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`.

```

6693 \def\bbl@tempa{\csname newif\endcsname&ifin@}
6694 \catcode`\&=4
6695 \ifx\in@\@undefined
6696 \def\in@#1#2{%
6697 \def\in@@##1##2##3\in@@{%
6698 \ifx\in@##2\in@false\else\in@true\fi}%
6699 \in@@#1\in@\in@@}
6700 \else
6701 \let\bbl@tempa\@empty
6702 \fi
6703 \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 T_EX 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).

```
6704 \def\@ifpackagewith#1#2#3#4{#3}
```

The L^AT_EX macro \ifl@aded checks whether a file was loaded. This functionality is not needed for plain T_EX but we need the macro to be defined as a no-op.

```
6705 \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 L^AT_EX 2_ε versions; just enough to make things work in plain T_EX environments.

```
6706 \ifx\@tempcnta\@undefined
6707   \csname newcount\endcsname\@tempcnta\relax
6708 \fi
6709 \ifx\@tempcntb\@undefined
6710   \csname newcount\endcsname\@tempcntb\relax
6711 \fi
```

To prevent wasting two counters in L^AT_EX 2.09 (because counters with the same name are allocated later by it) we reset the counter that holds the next free counter (\count10).

```
6712 \ifx\bye\@undefined
6713   \advance\count10 by -2\relax
6714 \fi
6715 \ifx\@ifnextchar\@undefined
6716   \def\@ifnextchar#1#2#3{%
6717     \let\reserved@d=#1%
6718     \def\reserved@a{#2}\def\reserved@b{#3}%
6719     \futurelet\@let@token\@ifnch}
6720 \def\@ifnch{%
6721   \ifx\@let@token\@sptoken
6722     \let\reserved@c\@xifnch
6723   \else
6724     \ifx\@let@token\reserved@d
6725       \let\reserved@c\reserved@a
6726     \else
6727       \let\reserved@c\reserved@b
6728     \fi
6729   \fi
6730   \reserved@c}
6731 \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
6732 \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
6733 \fi
6734 \def\@testopt#1#2{%
6735   \@ifnextchar[ {#1}{#1[#2]}}
6736 \def\@protected@testopt#1{%
6737   \ifx\protect\@typeset@protect
6738     \expandafter\@testopt
6739   \else
6740     \@x@protect#1%
6741   \fi}
6742 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
6743   #2\relax}\fi}
6744 \long\def\@iwhilenum#1{\ifnum #1\relax\expandafter\@iwhilenum
6745   \else\expandafter\@gobble\fi{#1}}
```

16.4 Encoding related macros

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

```
6746 \def\DeclareTextCommand{%
6747   \@dec@text@cmd\providecommand
6748 }
6749 \def\ProvideTextCommand{%
6750   \@dec@text@cmd\providecommand
6751 }
6752 \def\DeclareTextSymbol#1#2#3{%
6753   \@dec@text@cmd\chardef#1{#2}#3\relax
6754 }
6755 \def\@dec@text@cmd#1#2#3{%
6756   \expandafter\def\expandafter#2%
6757     \expandafter{%
6758       \csname#3-cmd\expandafter\endcsname
6759       \expandafter#2%
6760       \csname#3\string#2\endcsname
6761     }%
6762 %   \let\@ifdefinable\@rc@ifdefinable
6763   \expandafter#1\csname#3\string#2\endcsname
6764 }
6765 \def\@current@cmd#1{%
6766   \ifx\protect\@typeset@protect\else
6767     \noexpand#1\expandafter\@gobble
6768   \fi
6769 }
6770 \def\@changed@cmd#1#2{%
6771   \ifx\protect\@typeset@protect
6772     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
6773       \expandafter\ifx\csname ?\string#1\endcsname\relax
6774         \expandafter\def\csname ?\string#1\endcsname{%
6775           \@changed@x@err{#1}%
6776         }%
6777       \fi
6778       \global\expandafter\let
6779         \csname\cf@encoding \string#1\expandafter\endcsname
6780         \csname ?\string#1\endcsname
6781     \fi
6782     \csname\cf@encoding\string#1%
6783     \expandafter\endcsname
6784   \else
6785     \noexpand#1%
6786   \fi
6787 }
6788 \def\@changed@x@err#1{%
6789   \errhelp{Your command will be ignored, type <return> to proceed}%
6790   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
6791 \def\DeclareTextCommandDefault#1{%
6792   \DeclareTextCommand#1?%
6793 }
6794 \def\ProvideTextCommandDefault#1{%
6795   \ProvideTextCommand#1?%
6796 }
6797 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
6798 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
6799 \def\DeclareTextAccent#1#2#3{%
6800   \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
6801 }
```

```

6802 \def\DeclareTextCompositeCommand#1#2#3#4{%
6803   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
6804   \edef\reserved@b{\string##1}%
6805   \edef\reserved@c{%
6806     \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
6807   \ifx\reserved@b\reserved@c
6808     \expandafter\expandafter\expandafter\ifx
6809       \expandafter\@car\reserved@a\relax\relax\@nil
6810       \@text@composite
6811   \else
6812     \edef\reserved@b##1{%
6813       \def\expandafter\noexpand
6814         \csname#2\string#1\endcsname####1{%
6815         \noexpand\@text@composite
6816         \expandafter\noexpand\csname#2\string#1\endcsname
6817         ####1\noexpand\@empty\noexpand\@text@composite
6818         {##1}%
6819       }%
6820     }%
6821     \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
6822   \fi
6823   \expandafter\def\csname\expandafter\string\csname
6824     #2\endcsname\string#1-\string#3\endcsname{#4}
6825 \else
6826   \errhelp{Your command will be ignored, type <return> to proceed}%
6827   \errmessage{\string\DeclareTextCompositeCommand\space used on
6828     inappropriate command \protect#1}
6829 \fi
6830 }
6831 \def\@text@composite#1#2#3\@text@composite{%
6832   \expandafter\@text@composite@x
6833     \csname\string#1-\string#2\endcsname
6834 }
6835 \def\@text@composite@x#1#2{%
6836   \ifx#1\relax
6837     #2%
6838   \else
6839     #1%
6840   \fi
6841 }
6842 %
6843 \def\@strip@args#1:#2-#3\@strip@args{#2}
6844 \def\DeclareTextComposite#1#2#3#4{%
6845   \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
6846   \bgroup
6847     \lcode`\@=#4%
6848     \lowercase{%
6849   \egroup
6850     \reserved@a \@
6851   }%
6852 }
6853 %
6854 \def\UseTextSymbol#1#2{#2}
6855 \def\UseTextAccent#1#2#3{}
6856 \def\@use@text@encoding#1{}
6857 \def\DeclareTextSymbolDefault#1#2{%
6858   \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
6859 }
6860 \def\DeclareTextAccentDefault#1#2{%

```

```

6861 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
6862 }
6863 \def\cf@encoding{OT1}

```

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

```

6864 \DeclareTextAccent{"}{OT1}{127}
6865 \DeclareTextAccent{'}{OT1}{19}
6866 \DeclareTextAccent{^}{OT1}{94}
6867 \DeclareTextAccent`{OT1}{18}
6868 \DeclareTextAccent{~}{OT1}{126}

```

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

```

6869 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
6870 \DeclareTextSymbol{\textquotedblright}{OT1}{93}
6871 \DeclareTextSymbol{\textquoteleft}{OT1}{96}
6872 \DeclareTextSymbol{\textquoteright}{OT1}{97}
6873 \DeclareTextSymbol{\i}{OT1}{16}
6874 \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`.

```

6875 \ifx\scriptsize\@undefined
6876 \let\scriptsize\sevenrm
6877 \fi
6878 % End of code for plain
6879 <</Emulate LaTeX>>

```

A proxy file:

```

6880 <*plain>
6881 \input babel.def
6882 </plain>

```

17 Acknowledgements

I would like to thank all who volunteered as β -testers for their time. Michel Goossens supplied contributions for most of the other languages. Nico Poppelier helped polish the text of the documentation and supplied parts of the macros for the Dutch language. Paul Wackers and Werenfried Spit helped find and repair bugs. During the further development of the babel system I received much help from Bernd Raichle, for which I am grateful.

References

- [1] Huda Smitshuijzen Abifares, *Arabic Typography*, Saqi, 2001.
- [2] Johannes Braams, Victor Eijkhout and Nico Poppelier, *The development of national \LaTeX styles*, *TUGboat* 10 (1989) #3, p. 401–406.
- [3] Yannis Haralambous, *Fonts & Encodings*, O'Reilly, 2007.
- [4] Donald E. Knuth, *The \TeX book*, Addison-Wesley, 1986.
- [5] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [6] Leslie Lamport, *\LaTeX , A document preparation System*, Addison-Wesley, 1986.

- [7] Leslie Lamport, in: \TeX hax Digest, Volume 89, #13, 17 February 1989.
- [8] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [9] Hubert Partl, *German \TeX* , *TUGboat* 9 (1988) #1, p. 70–72.
- [10] Joachim Schrod, *International \LaTeX is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [11] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using \LaTeX* , Springer, 2002, p. 301–373.
- [12] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).