

Babel

Version 3.51.2218
2020/12/11

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	45
1.26	Languages supported by babel with ldf files	46
1.27	Unicode character properties in luatex	48
1.28	Tweaking some features	48
1.29	Tips, workarounds, known issues and notes	48
1.30	Current and future work	50
1.31	Tentative and experimental code	50
2	Loading languages with language.dat	50
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	53
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	56
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	61
6	locale directory	62
7	Tools	62
7.1	Multiple languages	67
7.2	The Package File (\LaTeX , babel.sty)	67
7.3	base	69
7.4	Conditional loading of shorthands	71
7.5	Cross referencing macros	73
7.6	Marks	76
7.7	Preventing clashes with other packages	77
7.7.1	ifthen	77
7.7.2	varioref	77
7.7.3	hhline	78
7.7.4	hyperref	78
7.7.5	fancyhdr	78
7.8	Encoding and fonts	79
7.9	Basic bidi support	81
7.10	Local Language Configuration	86
8	The kernel of Babel (babel.def, common)	90
8.1	Tools	90
9	Multiple languages	91
9.1	Selecting the language	94
9.2	Errors	102
9.3	Hooks	105
9.4	Setting up language files	107
9.5	Shorthands	109
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	132
9.12.3	Shorthands for quotation marks	133
9.12.4	Umlauts and tremas	134
9.13	Layout	136
9.14	Load engine specific macros	136
9.15	Creating and modifying languages	137
10	Adjusting the Babel bahavior	156
11	Loading hyphenation patterns	158
12	Font handling with fontspec	163

13	Hooks for XeTeX and LuaTeX	167
13.1	XeTeX	167
13.2	Layout	169
13.3	LuaTeX	171
13.4	Southeast Asian scripts	177
13.5	CJK line breaking	180
13.6	Automatic fonts and ids switching	181
13.7	Layout	192
13.8	Auto bidi with basic and basic-r	194
14	Data for CJK	205
15	The ‘nil’ language	205
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	207
16.4	Encoding related macros	211
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 $\LaTeX \geq 2018-04-01$ if the encoding is UTF-8.

PDFTEX

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

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

LUATEX/XETEX

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



```

\selectlanguage{vietnamese}

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

\end{document}

```

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` {*<language>*}

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

```
\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` [*<option-list>*]{*<language>*}{*<text>*}

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

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

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}` {*<language>*} ... `\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 `\knbbcode`, 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 `ngerman` 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 options to set `'` as a shorthand in case it is not done by default.
- activegrave** Same for ```.
- shorthands=** $\langle char \rangle \langle char \rangle \dots \mid \text{off}$
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 `{\langle option-name \rangle}{\langle code \rangle}`

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

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

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

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

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 lualatex also applies here. Some quick patterns can help, with something similar to:

```
\babelprovide[import,hyphenrules=+]{lao}
\babelpatterns[lao]{lᦺᦑ ᦺᦑ ᦺᦑ ᦺᦑ ᦺᦑ ᦺᦑ} % 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	cognian
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 imported 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\captionenglish{%  
  \renewcommand\contentsname{Foo}%  
}
```

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

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`.

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 an additional replacement type with the key `string`.

EXAMPLE 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 dictionaries into account). For example, you can use the `string` replacement to 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
\babelposthyphenation{russian-latin}{([sz])h} % Create rule
{
  { string = {1|sz|šž} },
  remove
}
```

In other words, it is a quite general tool. (A counterpart `\babelprehyphenation` is on the way.)

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` → `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

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.

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:

```
\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).
    فصحي العصر فصحي التراث

\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 `\textthe` must be defined to select the main language):

```

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

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 \LaTeX , 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 `\LaTeX2e` **New 3.19** .

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

EXAMPLE Typically, in an Arabic document you would need:

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

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

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

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

There are *three* R blocks and *two* L blocks, and the order is *RTL B and still ltr 1 ltr text RTL A*. This is by design to provide the proper 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 $\{\langle section\text{-}name\rangle\}$

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

New 3.17 Something like:

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

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

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

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

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

(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 T_EX parameters (#1, #2, #3), with the meaning given:

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

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

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

defaultcommands Used (locally) in `\StartBabelCommands`.

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

stopcommands Used to reset the above, if necessary.

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

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

afterextras Just after executing `\extras<language>`. For example, the following deactivates 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
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}

```

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

Then you preprocess it with `devnag <file>`, which creates `<file>.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).

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

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 `luahbtex` you may need `bidi.mirroring=off`. Use with care, because these options do not deactivate other related options (like `paragraph direction` with `bidi.text`).

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

(A recent version of inputenc is required.)

- For the hyphenation to work correctly, lccodes cannot change, because T_EX only takes into account the values when the paragraph is hyphenated, i.e., when it has been finished.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 T_EX, not of babel. Alternatively, you may use \usesorthands to activate ' and \defineshortand, or redefine \textquoteright (the latter is called by the non-ASCII right quote).
- \bibitem is out of sync with \selectlanguage in the .aux file. The reason is \bibitem uses \immediate (and others, in fact), while \selectlanguage doesn't. There is no known workaround.
- Babel does not take into account \normalsfcodes and (non-)French spacing is not always properly (un)set by languages. However, problems are unlikely to happen and therefore this part remains untouched in version 3.9 (but it is in the 'to do' list).
- Using a character mathematically active (ie, with math code "8000) as a shorthand can make T_EX enter in an infinite loop in some rare cases. (Another issue in the 'to do' list, although there is a partial solution.)

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

csquotes Logical markup for quotes.

iflang Tests correctly the current language.

hyphsubst Selects a different set of patterns for a language.

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

siunitx Typesetting of numbers and physical quantities.

biblatex Programmable bibliographies and citations.

bicaption Bilingual captions.

babelbib Multilingual bibliographies.

microtype Adjusts the typesetting according to some languages (kerning and spacing).
Ligatures can be disabled.

substitutefont Combines fonts in several encodings.

mkpattern Generates hyphenation patterns.

tracklang Tracks which languages have been requested.

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

zhspacing Spacing for CJK documents in xetex.

²⁰This explains why L^AT_EX assumes the lowercase mapping of T1 and does not provide a tool for multiple mappings. Unfortunately, \savingsphcodes is not a solution either, because lccodes for hyphenation are frozen in the format and cannot be changed.

1.30 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 \LaTeX internals. Calendars (Arabic, Persian, Indic, etc.) are under study.

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

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

`\babelprehyphenation`

New 3.44 Note it is tentative, but the current behavior for glyphs should be correct. It is similar to `\babelposthyphenation`, 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 (| is still reserved, but currently unused); (3) in the replacement, discretionaries are not accepted, only remove, , and string = ...

Currently it handles glyphs, not discretionaries or spaces (in particular, it will not catch the hyphen and you can't insert or remove spaces). Also, you are limited to substitutions as done by lua, although a future implementation may alternatively accept lpeg. Performance is still somewhat poor.

2 Loading languages with `language.dat`

\TeX and most engines based on it (pdf \TeX , xetex, ϵ - \TeX , the main exception being luatex) require hyphenation patterns to be preloaded when a format is created (eg, \LaTeX , Xe \LaTeX , pdf \LaTeX). babel provides a tool which has become standard in many distributions and based on a “configuration file” named `language.dat`. The exact way this file is used

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.

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
```

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.

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.

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

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

The following assumptions are made:

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

Some recommendations:

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

- 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.

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

²⁶But not removed, for backward compatibility.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.3 Skeleton

Here is the basic structure of an ldf file, with a language, a dialect and an attribute. Strings are best defined using the method explained in 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>

\bb{<language>}{<attrib>}{%
  \expandafter\addto\expandafter\extras<language>
  \expandafter{\extras<attrib><language>}%
  \let\captions<language>\captions<attrib><language>}

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

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

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

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

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

\EndBabelCommands

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

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

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

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


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

3.4 Support for active characters

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

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

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

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

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

3.5 Support for saving macro definitions

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

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

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

3.6 Support for extending macros

`\addto` The macro `\addto{⟨control sequence⟩}{⟨ \TeX code⟩}` can be used to extend the definition of a macro. The macro need not be defined (ie, it can be undefined or `\relax`). This macro can, for instance, be used in adding instructions to a macro like `\extrasenglish`. Be careful when using this macro, because depending on the case the assignment 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`.

²⁷This mechanism was introduced by Bernd Raichle.

3.7 Macros common to a number of languages

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

3.8 Encoding-dependent strings

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

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

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

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

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

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

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

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

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

The *category* is either captions, date or extras. You must stick to these three categories, even if no error is raised when using other name.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\monthinname{Jänner}

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

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

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

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

²⁸In future releases further categories may be added.

```

\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 $\langle date \rangle \langle language \rangle$ exists).

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

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

\EndBabelCommands Marks the end of the series of blocks.

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

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

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

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

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

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

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

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 $\langle map-list \rangle$ is a series of macros using the internal format of `@uclclist` (eg, `\bb\BB\cc\CC`). The mandatory

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

arguments take precedence over the optional one. This command, unlike `\SetString`, is executed always (even without strings), and it is intended for minor readjustments only. For example, as T1 is the default case mapping in \LaTeX , we 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` $\{ \langle to\text{-}lower\text{-}macros \rangle \}$

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

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

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

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

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

4 Changes

4.1 Changes in babel version 3.9

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

- `\select@language` did not set `\language`name. This meant the language in force when auxiliary files were loaded was the one used in, for example, shorthands – if the language was german, a `\select@language{spanish}` had no effect.
- `\foreignlanguage` and `otherlanguage*` messed up `\extras<language>`. Scripts, encodings and many other things were not switched correctly.
- The `:ENC` mechanism for hyphenation patterns used the encoding of the *previous* language, not that of the language being selected.
- `'` (with `activeacute`) had the original value when writing to an auxiliary file, and things like an infinite loop 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 \TeX 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 a uppercase letter. It can be just a letter (eg, babel.name.A, babel.name.B) or a name (eg, date.long.Nominative, date.long.Formal, but no language is currently using the latter). *Multi-letter* qualifiers are forward compatible in the sense they won’t conflict with new “global” keys (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=3.51.2218>>

2 <<date=2020/12/11>>

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

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

```

3 <<(*Basic macros)>> ≡
4 \bbl@trace{Basic macros}
5 \def\bbl@stripslash{\expandafter\@gobble\string}
6 \def\bbl@add#1#2{%
7   \bbl@ifunset{\bbl@stripslash#1}%
8   {\def#1{#2}}%
9   {\expandafter\def\expandafter#1\expandafter{#1#2}}
10 \def\bbl@xin@{\@expandtwoargs\in@}
11 \def\bbl@csarg#1#2{\expandafter#1\csname bbl@#2\endcsname}%
12 \def\bbl@cs#1{\csname bbl@#1\endcsname}
13 \def\bbl@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 Because the code that is used in the handling of active characters may need to look ahead,
\bbl@afterfi we take extra care to ‘throw’ it over the \else and \fi parts of an \if-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
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{%

```

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


```

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

```

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

```

`\bb@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\bb@ifblank#1{%
69 \bb@ifblank@i#1\@nil\@nil\@secondoftwo\@firstoftwo\@nil}
70 \long\def\bb@ifblank@i#1#2\@nil#3#4#5\@nil{#4}
71 \def\bb@ifset#1#2#3{%
72 \bb@ifunset{#1}{#3}{\bb@exp{\bb@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\bb@forkv#1#2{%
74 \def\bb@kvcmd##1##2##3{#2}%
75 \bb@kvnext#1,\@nil,}
76 \def\bb@kvnext#1,{%
77 \ifx\@nil#1\relax\else
78 \bb@ifblank{#1}{\bb@forkv@eq#1=\@empty=\@nil{#1}}%
79 \expandafter\bb@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 pdf \TeX , 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\bb1@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 ~~TeX~~ 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\bb@iniflag\@ne} % main -> +1
356 \DeclareOption{provide+=*}{\chardef\bb@iniflag\tw@} % add = 2
357 \DeclareOption{provide*=*}{\chardef\bb@iniflag\thr@@} % add + main
358 % A separate option
359 \let\bb@autoload@options\@empty
360 \DeclareOption{provide@=*}{\def\bb@autoload@options{import}}
361 % Don't use. Experimental. TODO.
362 \newif\ifbb@single
363 \DeclareOption{selectors=off}{\bb@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\bb@opt@shorthands\@nnil
366 \let\bb@opt@config\@nnil
367 \let\bb@opt@main\@nnil
368 \let\bb@opt@headfoot\@nnil
369 \let\bb@opt@layout\@nnil

```

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

```

370 \def\bb@tempa#1=#2\bb@tempa{%
371   \bb@csarg\ifx{opt@#1}\@nnil
372     \bb@csarg\edef{opt@#1}{#2}%
373   \else
374     \bb@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 `\bb@language@opts`, because they are language options.

```

381 \let\bb@language@opts\@empty
382 \DeclareOption*{%
383   \bb@xin@{\string=}{\CurrentOption}%
384   \ifin@
385     \expandafter\bb@tempa\CurrentOption\bb@tempa
386   \else
387     \bb@add@list\bb@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 `\bb@ifshorthand` is always true, and it is always false if `shorthands` is empty. Also, some code makes sense only with `shorthands=...`


```

390 \bbl@trace{Conditional loading of shorthands}
391 \def\bbl@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\bbl@sh@string
398 \fi}
399 \ifx\bbl@opt@shorthands\@nnil
400   \def\bbl@ifshorthand#1#2#3{#2}%
401 \else\ifx\bbl@opt@shorthands\@empty
402   \def\bbl@ifshorthand#1#2#3{#3}%
403 \else

```

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

```

404 \def\bbl@ifshorthand#1{%
405   \bbl@xin@{\string#1}{\bbl@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\bbl@opt@shorthands{%
412   \expandafter\bbl@sh@string\bbl@opt@shorthands\@empty}%

```

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

```

413 \bbl@ifshorthand{'}%
414   {\PassOptionsToPackage{activeacute}{babel}}{}
415 \bbl@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\bbl@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@activesfalse
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 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@activesfalse}
485 \bbl@redefineroobust\pageref#1{%
486 \@safe@activetrue\org@pageref{#1}\@safe@activesfalse}
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 \LaTeX 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 \LaTeX 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}%
548 \protect\bbl@restore@actives\the\toks@}}}%

```

`\markboth` The definition of `\markboth` is equivalent to that of `\markright`, except that we need two token registers. The documentclasses report and book define and set the headings for the page. While doing so they also store a copy of `\markboth` in `\@mkboth`. Therefore we need to check whether `\@mkboth` has already been set. If so we need to do that again with the new definition of `\markboth`. (As of Oct 2019, \LaTeX stores the definition in an intermediate macro, so it’s not necessary anymore, but it’s preserved for older versions.)

`\@mkboth`

```

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}{\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\bbl@temp@pref
585        \let\ref\bbl@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`
`\vrefpagemum` in order to prevent problems when an active character ends up in the argument of `\vref`.
`\Ref` The same needs to happen for `\vrefpagemum`.

```

591 \AtBeginDocument{%

```

```

592 \ifpackageloaded{varioref}{%
593   \bbl@redefine\@vpageref#1[#2]#3{%
594     \@safe@activetrue
595     \org@@vpageref{#1}[#2]#3}%
596   \@safe@activesfalse}%
597 \bbl@redefine\hrefpagenum#1#2{%
598   \@safe@activetrue
599   \org@hrefpagenum{#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 provides by \LaTeX .

```

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

`\ensureascii`

```

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@else
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 \TeX grouping.
- `luatex` can provide the most complete solution, as we can manipulate almost freely the node list, the generated lines, and so on, but bidi text does not work out of the box and some development is necessary. It also provides tools to properly set left-to-right and right-to-left page layouts. As `Lua \TeX -ja` shows, vertical typesetting is possible, too.

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,Manichaean,%
817 Manichaean,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\bbl@provide@dirs#1{%
822 \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts\bbl@rscripts}%
823 \ifin@
824 \global\bbl@csarg\chardef{wdir@#1}\@ne
825 \bbl@xin@{\csname bbl@sname@#1\endcsname}{\bbl@alscripts}%
826 \ifin@
827 \global\bbl@csarg\chardef{wdir@#1}\tw@ % useless in xetex
828 \fi
829 \else
830 \global\bbl@csarg\chardef{wdir@#1}\z@
831 \fi
832 \ifodd\bbl@engine
833 \bbl@csarg\ifcase{wdir@#1}%
834 \directlua{ Babel.locale_props[\the\localeid].textdir = 'l' }%
835 \or
836 \directlua{ Babel.locale_props[\the\localeid].textdir = 'r' }%

```

```

837 \or
838 \directlua{ Babel.locale_props[\the\localeid].texdir = 'al' }%
839 \fi
840 \fi}
841 \def\bbl@switchdir{%
842 \bbl@ifunset{\bbl@lsys@\language}{\bbl@provide@lsys{\language}}{}%
843 \bbl@ifunset{\bbl@wdir@\language}{\bbl@provide@dirs{\language}}{}%
844 \bbl@exp{\bbl@setdirs\bbl@cl{wdir}}%
845 \def\bbl@setdirs#1{% TODO - math
846 \ifcase\bbl@select@type % TODO - strictly, not the right test
847 \bbl@bodydir{#1}%
848 \bbl@pdir{#1}%
849 \fi
850 \bbl@texdir{#1}}
851 % TODO. Only if \bbl@bidimode > 0?:
852 \AddBabelHook{babel-bidi}{afterextras}{\bbl@switchdir}
853 \DisableBabelHook{babel-bidi}

```

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

```

854 \ifodd\bbl@engine % luatex=1
855 \chardef\bbl@thetexdir\z@
856 \chardef\bbl@thepardir\z@
857 \def\bbl@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\bbl@setluadir#1#2#3{% 1=text/par.. 2=\texdir.. 3=0 lr/1 rl
865 \ifcase#3\relax
866 \ifcase\bbl@getluadir{#1}\relax\else
867 #2 TLT\relax
868 \fi
869 \else
870 \ifcase\bbl@getluadir{#1}\relax
871 #2 TRT\relax
872 \fi
873 \fi}
874 \def\bbl@texdir#1{%
875 \bbl@setluadir{tex}\texdir{#1}%
876 \chardef\bbl@thetexdir#1\relax
877 \setattribute\bbl@attr@dir{\numexpr\bbl@thepardir*3+#1}}
878 \def\bbl@pdir#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\texdir\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@thetexdir\relax
888 \everyhbox{\texdir TLT\relax}%
889 \else
890 \everyhbox{\texdir 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\currentgrouptype\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\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@pdir#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@xeeverypar{%
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 `norsk.cfg` will be loaded when the language definition file `norsk.ldf` is loaded.

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

```

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 \LaTeX 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 \let\bbl@temp@lg\@empty
1057 \bbl@foreach\bbl@language@opts{%
1058   \ifcase\bbl@iniflag % Default
1059     \bbl@ifunset{ds@#1}%
1060     {\DeclareOption{#1}{\bbl@load@language{#1}}}%
1061     {}%
1062   \or % provide=*
1063     \@gobble % case 2 same as 1
1064   \or % provide+=*
1065     \bbl@ifunset{ds@#1}%
1066     {\IfFileExists{#1.ldf}{}%
1067       {\IfFileExists{babel-#1.tex}{\@namedef{ds@#1}}{}}}%
1068     {}%
1069     \bbl@ifunset{ds@#1}%
1070     {\def\bbl@tempc{#1}%
1071       \DeclareOption{#1}{%
1072         \ifnum\bbl@iniflag>\@ne
1073           \bbl@ldfinit
1074           \babelprovide[import]{#1}%
1075           \bbl@afterldf{}}%
1076         \else
1077           \bbl@load@language{#1}%
1078         \fi}%
1079     }%
1080   \or % provide*=*
1081     \def\bbl@tempc{#1}%
1082     \bbl@ifunset{ds@#1}%
1083     {\DeclareOption{#1}{%
1084       \bbl@ldfinit
1085       \babelprovide[import]{#1}%
1086       \bbl@afterldf{}}}%
1087     {}%
1088   \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.

```

1089 \let\bbl@tempb\@nnil
1090 \bbl@foreach\@classoptionslist{%
1091   \bbl@ifunset{ds@#1}%
1092     {\IfFileExists{#1.ldf}{}%
1093      {\IfFileExists{babel-#1.tex}{\@namedef{ds@#1}{}}}%
1094      {}}%
1095   \bbl@ifunset{ds@#1}%
1096     {\def\bbl@tempb{#1}%
1097      \DeclareOption{#1}{%
1098        \ifnum\bbl@iniflag>\@ne
1099          \bbl@ldfinit
1100          \babelprovide[import]{#1}%
1101          \bbl@afterldf}%
1102        \else
1103          \bbl@load@language{#1}%
1104        \fi}%
1105     {}}

```

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

```

1106 \ifnum\bbl@iniflag=\z@ \else
1107   \ifx\bbl@opt@main\@nnil
1108     \ifx\bbl@tempc\relax
1109       \let\bbl@opt@main\bbl@tempb
1110     \else
1111       \let\bbl@opt@main\bbl@tempc
1112     \fi
1113   \fi
1114 \fi
1115 \ifx\bbl@opt@main\@nnil \else
1116   \expandafter
1117   \let\expandafter\bbl@loadmain\csname ds@\bbl@opt@main\endcsname
1118   \expandafter\let\csname ds@\bbl@opt@main\endcsname\@empty
1119 \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):

```

1120 \def\AfterBabelLanguage#1{%
1121   \bbl@ifsamestring\CurrentOption{#1}{\global\bbl@add\bbl@afterlang}{}}
1122 \DeclareOption*{}
1123 \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`.

```

1124 \bbl@trace{Option 'main'}
1125 \ifx\bbl@opt@main\@nnil
1126   \edef\bbl@tempa{\@classoptionslist,\bbl@language@opts}
1127   \let\bbl@tempc\@empty
1128   \bbl@for\bbl@tempb\bbl@tempa{%
1129     \bbl@xin@{,\bbl@tempb,}{,\bbl@loaded,}%
1130     \ifin\edef\bbl@tempc{\bbl@tempb}\fi}
1131   \def\bbl@tempa#1,#2\@nnil{\def\bbl@tempb{#1}}
1132   \expandafter\bbl@tempa\bbl@loaded,\@nnil
1133   \ifx\bbl@tempb\bbl@tempc \else

```

```

1134 \bbl@warning{%
1135     Last declared language option is '\bbl@tempc',\%
1136     but the last processed one was '\bbl@tempb'.\%
1137     The main language cannot be set as both a global\%
1138     and a package option. Use 'main=\bbl@tempc' as\%
1139     option. Reported}%
1140 \fi
1141 \else
1142 \ifodd\bbl@iniflag % case 1,3
1143 \bbl@ldfinit
1144 \let\CurrentOption\bbl@opt@main
1145 \bbl@exp{\bbl@babelprovide[import,main]{\bbl@opt@main}}
1146 \bbl@afterldf{%
1147 \else % case 0,2
1148 \chardef\bbl@iniflag\z@ % Force ldf
1149 \expandafter\let\csname ds@\bbl@opt@main\endcsname\bbl@loadmain
1150 \ExecuteOptions{\bbl@opt@main}
1151 \DeclareOption*{%
1152 \ProcessOptions*
1153 \fi
1154 \fi
1155 \def\AfterBabelLanguage{%
1156 \bbl@error
1157 {Too late for \string\AfterBabelLanguage}%
1158 {Languages have been loaded, so I can do nothing}}
1159 \ifx\bbl@main@language\@undefined
1160 \bbl@info{%
1161     You haven't specified a language. I'll use 'nil'\%
1162     as the main language. Reported}
1163 \bbl@load@language{nil}
1164 \fi
1165 \</package>
1166 \*core>

```

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.

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 $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ users might want to use some of the features of the babel system too, care has to be taken that plain $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ 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 $\mathrm{T}_{\mathrm{E}}\mathrm{X}$ and $\mathrm{L}_{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$, some of it is for the $\mathrm{L}_{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$ 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

```

1167 \ifx\ldf@quit\@undefined\else
1168 \endinput\fi % Same line!
1169 \<<Make sure ProvidesFile is defined>>

```

```
1170 \ProvidesFile{babel.def}[\langle date \rangle \langle version \rangle] Babel common definitions]
```

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

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

```
1171 \ifx\AtBeginDocument\@undefined % TODO. change test.
1172 \langle Emulate LaTeX \rangle
1173 \def\language{english}%
1174 \let\bbl@opt@shorthands\@nnil
1175 \def\bbl@ifshorthand#1#2#3{#2}%
1176 \let\bbl@language@opts\@empty
1177 \ifx\babeloptionstrings\@undefined
1178   \let\bbl@opt@strings\@nnil
1179 \else
1180   \let\bbl@opt@strings\babeloptionstrings
1181 \fi
1182 \def\BabelStringsDefault{generic}
1183 \def\bbl@tempa{normal}
1184 \ifx\babeloptionmath\bbl@tempa
1185   \def\bbl@mathnormal{\noexpand\textormath}
1186 \fi
1187 \def\AfterBabelLanguage#1#2{}
1188 \ifx\BabelModifiers\@undefined\let\BabelModifiers\relax\fi
1189 \let\bbl@afterlang\relax
1190 \def\bbl@opt@safe{BR}
1191 \ifx\@uclclist\@undefined\let\@uclclist\@empty\fi
1192 \ifx\bbl@trace\@undefined\def\bbl@trace#1{}\fi
1193 \expandafter\newif\csname ifbbl@single\endcsname
1194 \chardef\bbl@bidimode\z@
1195 \fi
```

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

```
1196 \ifx\bbl@trace\@undefined
1197   \let\LdfInit\endinput
1198 \def\ProvidesLanguage#1{\endinput}
1199 \endinput\fi % Same line!
```

And continue.

9 Multiple languages

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

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.

```
1200 \langle Define core switching macros \rangle
```

`\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.

```
1201 \def\bbl@version{\langle version \rangle}
1202 \def\bbl@date{\langle date \rangle}
1203 \def\adddialect#1#2{%
1204   \global\chardef#1#2\relax
1205   \bbl@usehooks{adddialect}{\{#1\}\{#2\}}%
1206   \begingroup
```

```

1207 \count@#1\relax
1208 \def\bbl@elt##1##2##3##4{%
1209 \ifnum\count@=##2\relax
1210 \bbl@info{\string#1 = using hyphenrules for ##1\\%
1211 (\string\language\the\count@)}%
1212 \def\bbl@elt####1####2####3####4{%
1213 \fi}%
1214 \bbl@cs{languages}%
1215 \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.

```

1216 \def\bbl@fixname#1{%
1217 \begingroup
1218 \def\bbl@tempe{#1}%
1219 \edef\bbl@tempd{\noexpand\@ifundefined{\noexpand\bbl@tempe#1}}%
1220 \bbl@tempd
1221 {\lowercase\expandafter{\bbl@tempd}%
1222 {\uppercase\expandafter{\bbl@tempd}%
1223 \@empty
1224 {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1225 \uppercase\expandafter{\bbl@tempd}}}%
1226 {\edef\bbl@tempd{\def\noexpand#1{#1}}%
1227 \lowercase\expandafter{\bbl@tempd}}}%
1228 \@empty
1229 \edef\bbl@tempd{\endgroup\def\noexpand#1{#1}}%
1230 \bbl@tempd
1231 \bbl@exp{\@ifundefined{\bbl@usehooks{language}{\language}}{#1}}%
1232 \def\bbl@iflanguage#1{%
1233 \@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@bcpllookup either returns the found ini or it is \relax.

```

1234 \def\bbl@bcpcase#1#2#3#4\@#5{%
1235 \ifx\@empty#3%
1236 \uppercase{\def#5{#1#2}}%
1237 \else
1238 \uppercase{\def#5{#1}}%
1239 \lowercase{\edef#5{#5#2#3#4}}%
1240 \fi}
1241 \def\bbl@bcpllookup#1-#2-#3-#4\@{%
1242 \let\bbl@bcp\relax
1243 \lowercase{\def\bbl@tempa{#1}}%
1244 \ifx\@empty#2%
1245 \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1246 \else\ifx\@empty#3%
1247 \bbl@bcpcase#2\@empty\@empty\@{\bbl@tempb}
1248 \IfFileExists{babel-\bbl@tempa-\bbl@tempb.ini}%

```

```

1249     {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb}}%
1250     {}}%
1251     \ifx\bbl@bcp\relax
1252       \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1253     \fi
1254   \else
1255     \bbl@bcp#2\@empty\@empty\@empty\bbl@tempb
1256     \bbl@bcp#3\@empty\@empty\@empty\bbl@tempc
1257     \IfFileExists{babel-\bbl@tempa-\bbl@tempb-\bbl@tempc.ini}%
1258       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempb-\bbl@tempc}}%
1259       {}}%
1260     \ifx\bbl@bcp\relax
1261       \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1262       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
1263       {}}%
1264     \fi
1265     \ifx\bbl@bcp\relax
1266       \IfFileExists{babel-\bbl@tempa-\bbl@tempc.ini}%
1267       {\edef\bbl@bcp{\bbl@tempa-\bbl@tempc}}%
1268       {}}%
1269     \fi
1270     \ifx\bbl@bcp\relax
1271       \IfFileExists{babel-\bbl@tempa.ini}{\let\bbl@bcp\bbl@tempa}{}%
1272     \fi
1273   \fi\fi}
1274 \let\bbl@initoload\relax
1275 \def\bbl@provide@locale{%
1276   \ifx\babelprovide\undefined
1277     \bbl@error{For a language to be defined on the fly 'base'\\%
1278               is not enough, and the whole package must be\\%
1279               loaded. Either delete the 'base' option or\\%
1280               request the languages explicitly}%
1281     {See the manual for further details.}%
1282   \fi
1283 % TODO. Option to search if loaded, with \LocaleForEach
1284 \let\bbl@auxname\language % Still necessary. TODO
1285 \bbl@ifunset{bbl@bcp@map@\language}{}% Move uplevel??
1286 {\edef\language{\@nameuse{bbl@bcp@map@\language}}}%
1287 \ifbbl@bcpallowed
1288   \expandafter\ifx\csname date\language\endcsname\relax
1289     \expandafter
1290     \bbl@bcplookup\language-\@empty-\@empty-\@empty\@
1291     \ifx\bbl@bcp\relax\else % Returned by \bbl@bcplookup
1292       \edef\language{\bbl@bcp@prefix\bbl@bcp}%
1293       \edef\localename{\bbl@bcp@prefix\bbl@bcp}%
1294       \expandafter\ifx\csname date\language\endcsname\relax
1295         \let\bbl@initoload\bbl@bcp
1296         \bbl@exp{\@babelprovide[\bbl@autoload@bcptoptions]{\language}}%
1297         \let\bbl@initoload\relax
1298       \fi
1299       \bbl@csarg\xdef{bcp@map@\bbl@bcp}{\localename}%
1300     \fi
1301   \fi
1302 \fi
1303 \expandafter\ifx\csname date\language\endcsname\relax
1304   \IfFileExists{babel-\language.tex}%
1305   {\bbl@exp{\@babelprovide[\bbl@autoload@options]{\language}}}%
1306   {}}%
1307 \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.

```

1308 \def\iflanguage#1{%
1309   \bbl@iflanguage{#1}{%
1310     \ifnum\csname l@#1\endcsname=\language
1311       \expandafter\@firstoftwo
1312     \else
1313       \expandafter\@secondoftwo
1314     \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.

```

1315 \let\bbl@select@type\z@
1316 \edef\selectlanguage{%
1317   \noexpand\protect
1318   \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`.

```

1319 \ifx\@undefined\protect\let\protect\relax\fi

```

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

```

1320 \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.

```

1321 \def\bbl@language@stack{}

```

When using a stack we need a mechanism to push an element on the stack and to retrieve the information afterwards.

`\bbl@push@language` The stack is simply a list of languagenames, separated with a '+' sign; the push function can be simple:

`\bbl@pop@language`

```

1322 \def\bbl@push@language{%
1323   \ifx\language\@undefined\else
1324     \xdef\bbl@language@stack{\language+\bbl@language@stack}%
1325   \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`.

```
1326 \def\bbl@pop@lang#1+#2\@@{%
1327   \edef\language{#1}%
1328   \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).

```
1329 \let\bbl@ifrestoring\@secondoftwo
1330 \def\bbl@pop@language{%
1331   \expandafter\bbl@pop@lang\bbl@language@stack\@@
1332   \let\bbl@ifrestoring\@firstoftwo
1333   \expandafter\bbl@set@language\expandafter{\language}%
1334   \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).

```
1335 \chardef\localeid\z@
1336 \def\bbl@id@last{0} % No real need for a new counter
1337 \def\bbl@id@assign{%
1338   \bbl@ifunset{bbl@id@\language}%
1339   {\count@bbl@id@last\relax
1340    \advance\count@\@ne
1341    \bbl@csarg\chardef{id@\language}\count@
1342    \edef\bbl@id@last{\the\count@}%
1343    \ifcase\bbl@engine\or
1344      \directlua{
1345        Babel = Babel or {}
1346        Babel.locale_props = Babel.locale_props or {}
1347        Babel.locale_props[\bbl@id@last] = {}
1348        Babel.locale_props[\bbl@id@last].name = '\language'
1349      }%
1350    \fi}%
1351  }%
1352  \chardef\localeid\bbl@c{l{id@}}
```

The unprotected part of `\selectlanguage`.

```
1353 \expandafter\def\csname selectlanguage \endcsname#1{%
1354   \ifnum\bbl@hymapsel=\cclv\let\bbl@hymapsel\tw@fi
1355   \bbl@push@language
1356   \aftergroup\bbl@pop@language
1357   \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.

```

1358 \def\BabelContentsFiles{toc,lof,lot}
1359 \def\bbl@set@language#1{% from selectlanguage, pop@
1360 % The old buggy way. Preserved for compatibility.
1361 \edef\language{%
1362   \ifnum\escapechar=\expandafter`\string#1\@empty
1363   \else\string#1\@empty\fi}%
1364 \ifcat\relax\noexpand#1%
1365   \expandafter\ifx\csname date\language\endcsname\relax
1366   \edef\language{#1}%
1367   \let\locale\language
1368 \else
1369   \bbl@info{Using '\string\language' instead of 'language' is\\%
1370             deprecated. If what you want is to use a\\%
1371             macro containing the actual locale, make\\%
1372             sure it does not not match any language.\\%
1373             Reported}%
1374 %             I'll\\%
1375 %             try to fix '\string\locale', but I cannot promise\\%
1376 %             anything. Reported}%
1377   \ifx\scantokens\undefined
1378     \def\locale{??}%
1379   \else
1380     \scantokens\expandafter{\expandafter
1381       \def\expandafter\locale\expandafter{\language}}%
1382   \fi
1383 \fi
1384 \else
1385   \def\locale{#1}% This one has the correct catcodes
1386 \fi
1387 \select@language{\language}%
1388 % write to aux
1389 \expandafter\ifx\csname date\language\endcsname\relax\else
1390   \if@files
1391     \ifx\babel@aux\@gobbletwo\else % Set if single in the first, redundant
1392       \protected@write\@auxout{\string\babel@aux{\bbl@auxname}}}%
1393     \fi
1394     \bbl@usehooks{write}}}%
1395   \fi
1396 \fi}
1397 %
1398 \newif\ifbbl@bcpallowed
1399 \bbl@bcpallowedfalse
1400 \def\select@language#1{% from set@, babel@aux
1401 % set hymap
1402 \ifnum\bbl@hymapset=\@cclv\chardef\bbl@hymapset4\relax\fi
1403 % set name
1404 \edef\language{#1}%
1405 \bbl@fixname\language
1406 % TODO. name@map must be here?
1407 \bbl@provide@locale
1408 \bbl@iflanguage\language{%
1409   \expandafter\ifx\csname date\language\endcsname\relax

```

```

1410 \bbl@error
1411 {Unknown language '\language'. Either you have\\%
1412 misspelled its name, it has not been installed,\\%
1413 or you requested it in a previous run. Fix its name,\\%
1414 install it or just rerun the file, respectively. In\\%
1415 some cases, you may need to remove the aux file}%
1416 {You may proceed, but expect wrong results}%
1417 \else
1418 % set type
1419 \let\bbl@select@type\z@
1420 \expandafter\bbl@switch\expandafter{\language}%
1421 \fi}}
1422 \def\babel@aux#1#2{% TODO. See how to avoid undefined nil's
1423 \select@language{#1}%
1424 \bbl@foreach\BabelContentsFiles{%
1425 \@writefile{##1}{\babel@toc{#1}{#2}}}% %% TODO - ok in plain?
1426 \def\babel@toc#1#2{%
1427 \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.

```

1428 \newif\ifbbl@usedategroup
1429 \def\bbl@switch#1{% from select@, foreign@
1430 % make sure there is info for the language if so requested
1431 \bbl@ensureinfo{#1}%
1432 % restore
1433 \originalTeX
1434 \expandafter\def\expandafter\originalTeX\expandafter{%
1435 \csname noextras#1\endcsname
1436 \let\originalTeX\@empty
1437 \babel@beginsave}%
1438 \bbl@usehooks{afterreset}}}%
1439 \languageshorthands{none}%
1440 % set the locale id
1441 \bbl@id@assign
1442 % switch captions, date
1443 % No text is supposed to be added here, so we remove any
1444 % spurious spaces.
1445 \bbl@bsphack
1446 \ifcase\bbl@select@type
1447 \csname captions#1\endcsname\relax
1448 \csname date#1\endcsname\relax
1449 \else
1450 \bbl@xin@{,captions,},{, \bbl@select@opts,}%
1451 \ifin@

```

```

1452     \csname captions#1\endcsname\relax
1453     \fi
1454     \bbl@xin@{,date,}{, \bbl@select@opts,}%
1455     \ifin@ % if \foreign... within \<lang>date
1456     \csname date#1\endcsname\relax
1457     \fi
1458     \fi
1459     \bbl@esphack
1460     % switch extras
1461     \bbl@usehooks{beforeextras}{}%
1462     \csname extras#1\endcsname\relax
1463     \bbl@usehooks{afterextras}{}%
1464     % > babel-ensure
1465     % > babel-sh-<short>
1466     % > babel-bidi
1467     % > babel-fontspec
1468     % hyphenation - case mapping
1469     \ifcase\bbl@opt@hyphenmap\or
1470     \def\BabelLower##1##2{\lccode##1=##2\relax}%
1471     \ifnum\bbl@hymapsel>4\else
1472     \csname\language\language @bbl@hyphenmap\endcsname
1473     \fi
1474     \chardef\bbl@opt@hyphenmap\z@
1475     \else
1476     \ifnum\bbl@hymapsel>\bbl@opt@hyphenmap\else
1477     \csname\language\language @bbl@hyphenmap\endcsname
1478     \fi
1479     \fi
1480     \let\bbl@hymapsel\@cclv
1481     % hyphenation - select patterns
1482     \bbl@patterns{#1}%
1483     % hyphenation - allow stretching with babelnohyphens
1484     \ifnum\language=\l@babelnohyphens
1485     \babel@savevariable\emergencystretch
1486     \emergencystretch\maxdimen
1487     \babel@savevariable\hbadness
1488     \hbadness\@M
1489     \fi
1490     % hyphenation - mins
1491     \babel@savevariable\lefthyphenmin
1492     \babel@savevariable\righthyphenmin
1493     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1494     \set@hyphenmins\tw@\thr@\relax
1495     \else
1496     \expandafter\expandafter\expandafter\set@hyphenmins
1497     \csname #1hyphenmins\endcsname\relax
1498     \fi}

```

otherlanguage The other language 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.

```

1499 \long\def\otherlanguage#1{%
1500 \ifnum\bbl@hymapsel=\@cclv\let\bbl@hymapsel\thr@\fi
1501 \csname selectlanguage \endcsname{#1}%
1502 \ignorespaces}

```

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

```
1503 \long\def\endotherlanguage{%
1504   \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`.

```
1505 \expandafter\def\csname otherlanguage*\endcsname{%
1506   \@ifnextchar[\bbl@otherlanguage@s{\bbl@otherlanguage@s[]}}
1507 \def\bbl@otherlanguage@s[#1]#2{%
1508   \ifnum\bbl@hymapsel=\@ccclv\chardef\bbl@hymapsel4\relax\fi
1509   \def\bbl@select@opts{#1}%
1510   \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”.

```
1511 \expandafter\let\csname endotherlanguage*\endcsname\relax
```

`\foreignlanguage` The `\foreignlanguage` command is another substitute for the `\selectlanguage` command. This command takes two arguments, the first argument is the name of the language to use for typesetting the text specified in the second argument. Unlike `\selectlanguage` this command doesn’t switch *everything*, it only switches the hyphenation rules and the extra definitions for the language specified. It does this within a group and assumes the `\extras⟨lang⟩` command doesn’t make any `\global` changes. The coding is very similar to part of `\selectlanguage`. `\bbl@beforeforeign` is a trick to fix a bug in bidi texts. `\foreignlanguage` is supposed to be a ‘text’ command, and therefore it must emit a `\leavevmode`, but it does not, and therefore the indent is placed on the opposite margin. For backward compatibility, however, it is done only if a right-to-left script is requested; otherwise, it is no-op. (3.11) `\foreignlanguage*` is a temporary, experimental macro for a few lines with a different script direction, while preserving the paragraph format (thank the braces around `\par`, things like `\hangindent` are not reset). Do not use it in production, because its semantics and its syntax may change (and very likely will, or even it could be removed altogether). Currently it enters in `vmode` and then selects the language (which in turn sets the paragraph direction). (3.11) Also experimental are the hook `foreign` and `foreign*`. With them you can redefine `\BabelText` which by default does nothing. Its behavior is not well defined yet. So, use it in horizontal mode only if you do not want surprises. In other words, at the beginning of a paragraph `\foreignlanguage` enters into `hmode` with the surrounding `lang`, and with `\foreignlanguage*` with the new `lang`.

```
1512 \providecommand\bbl@beforeforeign{}
1513 \edef\foreignlanguage{%
1514   \noexpand\protect
1515   \expandafter\noexpand\csname foreignlanguage \endcsname}
1516 \expandafter\def\csname foreignlanguage \endcsname{%
1517   \@ifstar\bbl@foreign@s\bbl@foreign@x}
1518 \providecommand\bbl@foreign@x[3][]{%
1519   \begingroup
1520     \def\bbl@select@opts{#1}%
1521     \let\BabelText\@firstofone
1522     \bbl@beforeforeign
1523     \foreign@language{#2}%
1524     \bbl@usehooks{foreign}{}}%
```

```

1525 \BabelText{#3}% Now in horizontal mode!
1526 \endgroup}
1527 \def\bbbl@foreign@s#1#2{% TODO - \shapemode, \setpar, ?\@@par
1528 \begingroup
1529 {\par}%
1530 \let\BabelText\@firstofone
1531 \foreign@language{#1}%
1532 \bbbl@usehooks{foreign*}{}%
1533 \bbbl@dirparastext
1534 \BabelText{#2}% Still in vertical mode!
1535 {\par}%
1536 \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 `bbbl@switch`.

```

1537 \def\foreign@language#1{%
1538 % set name
1539 \edef\language#1}%
1540 \ifbbbl@usedategroup
1541 \bbbl@add\bbbl@select@opts{,date,}%
1542 \bbbl@usedategroupfalse
1543 \fi
1544 \bbbl@fixname\language
1545 % TODO. name@map here?
1546 \bbbl@provide@locale
1547 \bbbl@iflanguage\language{%
1548 \expandafter\ifx\csname date\language\endcsname\relax
1549 \bbbl@warning % TODO - why a warning, not an error?
1550 {Unknown language `#1'. Either you have\\%
1551 misspelled its name, it has not been installed,\\%
1552 or you requested it in a previous run. Fix its name,\\%
1553 install it or just rerun the file, respectively. In\\%
1554 some cases, you may need to remove the aux file.\\%
1555 I'll proceed, but expect wrong results.\\%
1556 Reported}%
1557 \fi
1558 % set type
1559 \let\bbbl@select@type\@ne
1560 \expandafter\bbbl@switch\expandafter{\language}}

```

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

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

```

1561 \let\bbbl@hyphlist\@empty
1562 \let\bbbl@hyphenation@relax
1563 \let\bbbl@pttnlist\@empty
1564 \let\bbbl@patterns@relax
1565 \let\bbbl@hymapsel=\@cclv
1566 \def\bbbl@patterns#1{%
1567 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax

```

```

1568     \csname l@#1\endcsname
1569     \edef\bbl@tempa{#1}%
1570     \else
1571     \csname l@#1:\f@encoding\endcsname
1572     \edef\bbl@tempa{#1:\f@encoding}%
1573     \fi
1574 \@@expandtwoargs\bbl@usehooks{patterns}{#{1}}{\bbl@tempa}}%
1575 % > luatex
1576 \@@ifundefined{bbl@hyphenation@}{% Can be \relax!
1577     \begingroup
1578     \bbl@xin@{,\number\language,}{,\bbl@hyphlist}%
1579     \ifin@else
1580     \@@expandtwoargs\bbl@usehooks{hyphenation}{#{1}}{\bbl@tempa}}%
1581     \hyphenation{%
1582     \bbl@hyphenation@
1583     \@@ifundefined{bbl@hyphenation@#1}%
1584     \empty
1585     {\space\csname bbl@hyphenation@#1\endcsname}}%
1586     \xdef\bbl@hyphlist{\bbl@hyphlist\number\language,}%
1587     \fi
1588 \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*.

```

1589 \def\hyphenrules#1{%
1590     \edef\bbl@tempf{#1}%
1591     \bbl@fixname\bbl@tempf
1592     \bbl@iflanguage\bbl@tempf{%
1593     \expandafter\bbl@patterns\expandafter{\bbl@tempf}%
1594     \ifx\languageshortands\@undefined\else
1595     \languageshortands{none}%
1596     \fi
1597     \expandafter\ifx\csname\bbl@tempf hyphenmins\endcsname\relax
1598     \set@hyphenmins\tw@\thr@@\relax
1599     \else
1600     \expandafter\expandafter\expandafter\set@hyphenmins
1601     \csname\bbl@tempf hyphenmins\endcsname\relax
1602     \fi}}
1603 \let\endhyphenrules\empty

```

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

```

1604 \def\providehyphenmins#1#2{%
1605     \expandafter\ifx\csname #1hyphenmins\endcsname\relax
1606     \@namedef{#1hyphenmins}{#2}%
1607     \fi}

```

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

```

1608 \def\set@hyphenmins#1#2{%
1609     \lefthyphenmin#1\relax
1610     \righthyphenmin#2\relax}

```

\ProvidesLanguage The identification code for each file is something that was introduced in $\text{\LaTeX 2}_{\epsilon}$. 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.

```

1611 \ifx\ProvidesFile\@undefined
1612 \def\ProvidesLanguage#1[#2 #3 #4]{%
1613   \wlog{Language: #1 #4 #3 <#2>}%
1614 }
1615 \else
1616 \def\ProvidesLanguage#1{%
1617   \begingroup
1618   \catcode\ 10 %
1619   \@makeother\/%
1620   \ifnextchar[%]
1621     {\@provideslanguage{#1}}{\@provideslanguage{#1}[]}%
1622 \def\@provideslanguage#1[#2]{%
1623   \wlog{Language: #1 #2}%
1624   \expandafter\xdef\csname ver@#1.ldf\endcsname{#2}%
1625   \endgroup}%
1626 \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`.

```

1627 \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.

```

1628 \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’:

```

1629 \providecommand\setlocale{%
1630   \bbl@error
1631   {Not yet available}%
1632   {Find an armchair, sit down and wait}}
1633 \let\uselocale\setlocale
1634 \let\locale\setlocale
1635 \let\selectlocale\setlocale
1636 \let\localename\setlocale
1637 \let\textlocale\setlocale
1638 \let\textlanguage\setlocale
1639 \let\languagetext\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 $\text{\LaTeX 2}_{\epsilon}$, so we can safely use its error handling interface. Otherwise we’ll have to ‘keep it simple’.
Infos are not written to the console, but on the other hand many people think warnings are errors, so a further message type is defined: an important info which is sent to the console.

```

1640 \edef\bbl@nulllanguage{\string\language=0}

```

```

1641 \ifx\PackageError\@undefined % TODO. Move to Plain
1642 \def\bbl@error#1#2{%
1643   \begingroup
1644     \newlinechar=`^^J
1645     \def\{^^J(babel) }%
1646     \errhelp{#2}\errmessage{\{#1}%
1647   \endgroup}
1648 \def\bbl@warning#1{%
1649   \begingroup
1650     \newlinechar=`^^J
1651     \def\{^^J(babel) }%
1652     \message{\{#1}%
1653   \endgroup}
1654 \let\bbl@infowarn\bbl@warning
1655 \def\bbl@info#1{%
1656   \begingroup
1657     \newlinechar=`^^J
1658     \def\{^^J}%
1659     \wlog{#1}%
1660   \endgroup}
1661 \fi
1662 \def\bbl@nocaption{\protect\bbl@nocaption@i}
1663 \def\bbl@nocaption@i#1#2{% 1: text to be printed 2: caption macro \langXname
1664   \global\@namedef{#2}{\textbf{?#1?}}%
1665   \@nameuse{#2}%
1666   \bbl@warning{%
1667     \@backslashchar#2 not set. Please, define it\\%
1668     after the language has been loaded (typically\\%
1669     in the preamble) with something like:\\%
1670     \string\renewcommand\@backslashchar#2{..}\\%
1671     Reported}}
1672 \def\bbl@tentative{\protect\bbl@tentative@i}
1673 \def\bbl@tentative@i#1{%
1674   \bbl@warning{%
1675     Some functions for '#1' are tentative.\\%
1676     They might not work as expected and their behavior\\%
1677     could change in the future.\\%
1678     Reported}}
1679 \def\@nolanerr#1{%
1680   \bbl@error
1681     {You haven't defined the language #1\space yet.\\%
1682     Perhaps you misspelled it or your installation\\%
1683     is not complete}%
1684     {Your command will be ignored, type <return> to proceed}}
1685 \def\@nopatterns#1{%
1686   \bbl@warning
1687     {No hyphenation patterns were preloaded for\\%
1688     the language `#1' into the format.\\%
1689     Please, configure your TeX system to add them and\\%
1690     rebuild the format. Now I will use the patterns\\%
1691     preloaded for \bbl@nulllanguage\space instead}}
1692 \let\bbl@usehooks\@gobbletwo
1693 \ifx\bbl@onlyswitch\@empty\endinput\fi
1694 % Here ended switch.def

Here ended switch.def.

1695 \ifx\directlua\@undefined\else
1696   \ifx\bbl@luapatterns\@undefined
1697     \input luababel.def

```



```

1698 \fi
1699 \fi
1700 <<Basic macros>>
1701 \bbl@trace{Compatibility with language.def}
1702 \ifx\bbl@languages\@undefined
1703 \ifx\directlua\@undefined
1704 \openin1 = language.def % TODO. Remove hardcoded number
1705 \ifeof1
1706 \closein1
1707 \message{I couldn't find the file language.def}
1708 \else
1709 \closein1
1710 \begingroup
1711 \def\addlanguage#1#2#3#4#5{%
1712 \expandafter\ifx\csname lang@#1\endcsname\relax\else
1713 \global\expandafter\let\csname l@#1\expandafter\endcsname
1714 \csname lang@#1\endcsname
1715 \fi}%
1716 \def\uselanguage#1{%
1717 \input language.def
1718 \endgroup
1719 \fi
1720 \fi
1721 \chardef\l@english\z@
1722 \fi

```

\addto It takes two arguments, a *<control sequence>* and TeX-code to be added to the *<control sequence>*.

If the *<control sequence>* has not been defined before it is defined now. The control sequence could also expand to `\relax`, in which case a circular definition results. The net result is a stack overflow. Note there is an inconsistency, because the assignment in the last branch is global.

```

1723 \def\addto#1#2{%
1724 \ifx#1\@undefined
1725 \def#1{#2}%
1726 \else
1727 \ifx#1\relax
1728 \def#1{#2}%
1729 \else
1730 {\toks@\expandafter{#1#2}%
1731 \xdef#1{\the\toks@}}%
1732 \fi
1733 \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?

```

1734 \def\bbl@withactive#1#2{%
1735 \begingroup
1736 \lccode`~=`#2\relax
1737 \lowercase{\endgroup#1~}}

```

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

(they have changed in the past). A macro named `\macro` will be saved new control sequences named `\org@macro`.

```
1738 \def\bbl@redefine#1{%
1739   \edef\bbl@tempa{\bbl@stripslash#1}%
1740   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1741   \expandafter\def\csname\bbl@tempa\endcsname}
1742 \@onlypreamble\bbl@redefine
```

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

```
1743 \def\bbl@redefine@long#1{%
1744   \edef\bbl@tempa{\bbl@stripslash#1}%
1745   \expandafter\let\csname org@\bbl@tempa\endcsname#1%
1746   \expandafter\long\expandafter\def\csname\bbl@tempa\endcsname}
1747 \@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_`.

```
1748 \def\bbl@redefineroobust#1{%
1749   \edef\bbl@tempa{\bbl@stripslash#1}%
1750   \bbl@ifunset{\bbl@tempa\space}%
1751   {\expandafter\let\csname org@\bbl@tempa\endcsname#1%
1752     \bbl@exp{\def\#1{\protect\<\bbl@tempa\space>}}}%
1753   {\bbl@exp{\let\<org@\bbl@tempa>\<\bbl@tempa\space>}}}%
1754   \@namedef{\bbl@tempa\space}}
1755 \@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.

```
1756 \bbl@trace{Hooks}
1757 \newcommand\AddBabelHook[3][{}]{%
1758   \bbl@ifunset{bbl@hk@#2}{\EnableBabelHook{#2}}{}%
1759   \def\bbl@tempa##1,#3=##2,##3\@empty{\def\bbl@tempb{##2}}%
1760   \expandafter\bbl@tempa\bbl@evargs,#3=,\@empty
1761   \bbl@ifunset{bbl@ev@#2@#3@#1}%
1762     {\bbl@csarg\bbl@add{ev@#3@#1}{\bbl@elth{#2}}}%
1763     {\bbl@csarg\let{ev@#2@#3@#1}\relax}%
1764   \bbl@csarg\newcommand{ev@#2@#3@#1}{\bbl@tempb}}
1765 \newcommand\EnableBabelHook[1]{\bbl@csarg\let{hk@#1}\@firstofone}
1766 \newcommand\DisableBabelHook[1]{\bbl@csarg\let{hk@#1}\@gobble}
1767 \def\bbl@usehooks#1#2{%
1768   \def\bbl@elth##1{%
1769     \bbl@cs{hk@##1}{\bbl@cs{ev@##1@#1}{#2}}%
1770     \bbl@cs{ev@#1@}%
1771     \ifx\language\@undefined\else % Test required for Plain (?)
1772       \def\bbl@elth##1{%
1773         \bbl@cs{hk@##1}{\bbl@cl{ev@##1@#1}{#2}}%
1774         \bbl@cl{ev@#1}%
1775       \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).

```

1776 \def\bbl@evargs{% <- don't delete this comma
1777   everylanguage=1,loadkernel=1,loadpatterns=1,loadexceptions=1,%
1778   adddialect=2,patterns=2,defaultcommands=0,encodedcommands=2,write=0,%
1779   beforeextras=0,afterextras=0,stopcommands=0,stringprocess=0,%
1780   hyphenation=2,initiateactive=3,afterreset=0,foreign=0,foreign*=0,%
1781   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.

```

1782 \bbl@trace{Defining babelensure}
1783 \newcommand\babelensure[2][{}% TODO - revise test files
1784   \AddBabelHook{babel-ensure}{afterextras}{%
1785     \ifcase\bbl@select@type
1786       \bbl@cl{e}%
1787       \fi}%
1788   \begingroup
1789     \let\bbl@ens@include\@empty
1790     \let\bbl@ens@exclude\@empty
1791     \def\bbl@ens@fontenc{\relax}%
1792     \def\bbl@tempb##1{%
1793       \ifx\@empty##1\else\noexpand##1\expandafter\bbl@tempb\fi}%
1794     \edef\bbl@tempa{\bbl@tempb#1\@empty}%
1795     \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ens@##1}{##2}}%
1796     \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
1797     \def\bbl@tempc{\bbl@ensure}%
1798     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1799       \expandafter{\bbl@ens@include}}%
1800     \expandafter\bbl@add\expandafter\bbl@tempc\expandafter{%
1801       \expandafter{\bbl@ens@exclude}}%
1802     \toks@\expandafter{\bbl@tempc}%
1803     \bbl@exp{%
1804   \endgroup
1805   \def\<bbl@e@#2>{\the\toks@{\bbl@ens@fontenc}}}%
1806 \def\bbl@ensure#1#2#3{% 1: include 2: exclude 3: fontenc
1807   \def\bbl@tempb##1{% elt for (excluding) \bbl@captionslist list
1808     \ifx##1\undefined % 3.32 - Don't assume the macro exists
1809       \edef##1{\noexpand\bbl@nocaption
1810         {\bbl@stripslash##1}{\language\bbl@stripslash##1}}%
1811     \fi
1812     \ifx##1\@empty\else
1813       \in@{##1}{#2}%
1814       \ifin\@else
1815         \bbl@ifunset{\bbl@ensure@\language}%
1816         {\bbl@exp{%
1817           \\\DeclareRobustCommand\bbl@ensure@\language>[1]{%

```

```

1818          \\\foreignlanguage{\language}%
1819          {\ifx\relax#3\else
1820            \\\fontencoding{#3}\selectfont
1821            \fi
1822            #####1}}}%
1823      {}%
1824      \toks@\expandafter{##1}%
1825      \edef##1{%
1826        \bbl@csarg\noexpand{ensure@\language}%
1827        {\the\toks@}}%
1828      \fi
1829      \expandafter\bbl@tempb
1830      \fi}%
1831      \expandafter\bbl@tempb\bbl@captionslist\today\@empty
1832      \def\bbl@tempa##1{% elt for include list
1833        \ifx##1\@empty\else
1834          \bbl@csarg\in@{ensure@\language\expandafter}\expandafter{##1}%
1835          \ifin@ \else
1836            \bbl@tempb##1\@empty
1837          \fi
1838          \expandafter\bbl@tempa
1839        \fi}%
1840      \bbl@tempa#1\@empty}
1841      \def\bbl@captionslist{%
1842        \prefacename\refname\abstractname\bibname\chaptername\appendixname
1843        \contentsname\listfigurename\listtablename\indexname\figurename
1844        \tablename\partname\encname\ccname\headtoname\pagename\seename
1845        \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`.

```

1846 \bbl@trace{Macros for setting language files up}
1847 \def\bbl@ldfinit{% TODO. Merge into the next macro? Unused elsewhere
1848   \let\bbl@screset\@empty
1849   \let\BabelStrings\bbl@opt@string
1850   \let\BabelOptions\@empty
1851   \let\BabelLanguages\relax

```

```

1852 \ifx\originalTeX\@undefined
1853   \let\originalTeX\@empty
1854 \else
1855   \originalTeX
1856 \fi}
1857 \def\LdfInit#1#2{%
1858   \chardef\atcatcode=\catcode`\@
1859   \catcode`\@=11\relax
1860   \chardef\eqcatcode=\catcode`\=
1861   \catcode`\==12\relax
1862   \expandafter\if\expandafter\@backslashchar
1863     \expandafter\@car\string#2\@nil
1864   \ifx#2\@undefined\else
1865     \ldf@quit{#1}%
1866   \fi
1867 \else
1868   \expandafter\ifx\csname#2\endcsname\relax\else
1869     \ldf@quit{#1}%
1870   \fi
1871 \fi
1872 \bbl@ldfinit}

```

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

```

1873 \def\ldf@quit#1{%
1874   \expandafter\main@language\expandafter{#1}%
1875   \catcode`\@=\atcatcode \let\atcatcode\relax
1876   \catcode`\==\eqcatcode \let\eqcatcode\relax
1877   \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.

```

1878 \def\bbl@afterldf#1{% TODO. Merge into the next macro? Unused elsewhere
1879   \bbl@afterlang
1880   \let\bbl@afterlang\relax
1881   \let\BabelModifiers\relax
1882   \let\bbl@screset\relax}%
1883 \def\ldf@finish#1{%
1884   \ifx\loadlocalcfg\@undefined\else % For LaTeX 209
1885     \loadlocalcfg{#1}%
1886   \fi
1887   \bbl@afterldf{#1}%
1888   \expandafter\main@language\expandafter{#1}%
1889   \catcode`\@=\atcatcode \let\atcatcode\relax
1890   \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 .

```

1891 \@onlypreamble\LdfInit
1892 \@onlypreamble\ldf@quit
1893 \@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.

```

1894 \def\main@language#1{%
1895   \def\bbl@main@language{#1}%
1896   \let\language\main@language % TODO. Set localename
1897   \bbl@id@assign
1898   \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`.

```

1899 \def\bbl@beforestart{%
1900   \bbl@usehooks{beforestart}{}%
1901   \global\let\bbl@beforestart\relax}
1902 \AtBeginDocument{%
1903   \@nameuse{bbl@beforestart}%
1904   \if@files
1905     \providecommand\babel@aux[2]{%
1906       \immediate\write\@mainaux{%
1907         \string\providecommand\string\babel@aux[2]{}}%
1908       \immediate\write\@mainaux{\string\@nameuse{bbl@beforestart}}%
1909     }
1910     \expandafter\selectlanguage\expandafter{\bbl@main@language}%
1911     \ifbbl@single % must go after the line above.
1912       \renewcommand\selectlanguage[1]{%
1913         \renewcommand\foreignlanguage[2]{#2}%
1914         \global\let\babel@aux\@gobbletwo % Also as flag
1915       }
1916     \ifcase\bbl@engine\or\pagedir\bodydir\fi} % TODO - a better place

```

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

```

1917 \def\select@language@x#1{%
1918   \ifcase\bbl@select@type
1919     \bbl@ifsamestring\language{#1}{\select@language{#1}}%
1920   \else
1921     \select@language{#1}%
1922   \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.

```

1923 \bbl@trace{Shorhands}
1924 \def\bbl@add@special#1{% 1:a macro like "\", \?, etc.
1925   \bbl@add\dospecials{\do#1}% test @sanitize = \relax, for back. compat.
1926   \bbl@ifunset{@sanitize}{\bbl@add\@sanitize{\@makeother#1}}%
1927   \if\nfss@catcodes\undefined\else % TODO - same for above
1928     \begingroup
1929       \catcode`#1\active
1930       \nfss@catcodes
1931       \ifnum\catcode`#1=\active
1932         \endgroup
1933       \bbl@add\nfss@catcodes{\@makeother#1}%

```

```

1934     \else
1935     \endgroup
1936     \fi
1937 \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.

```

1938 \def\bbl@remove@special#1{%
1939   \begingroup
1940   \def\x##1##2{\ifnum`#1=`##2\noexpand\@empty
1941     \else\noexpand##1\noexpand##2\fi}%
1942   \def\do{\x\do}%
1943   \def\@makeother{\x\@makeother}%
1944   \edef\x{\endgroup
1945     \def\noexpand\dospecials{\dospecials}%
1946     \expandafter\ifx\csname @sanitize\endcsname\relax\else
1947       \def\noexpand\@sanitize{\@sanitize}%
1948     \fi}%
1949   \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).

```

1950 \def\bbl@active@def#1#2#3#4{%
1951   \@namedef{#3#1}{%
1952     \expandafter\ifx\csname#2@sh@#1\endcsname\relax
1953       \bbl@afterelse\bbl@sh@select#2#1{#3@arg#1}{#4#1}%
1954     \else
1955       \bbl@afterfi\csname#2@sh@#1\endcsname
1956     \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.

```

1957   \long\@namedef{#3@arg#1}##1{%
1958     \expandafter\ifx\csname#2@sh@#1@string##1\endcsname\relax
1959       \bbl@afterelse\csname#4#1\endcsname##1%
1960     \else
1961       \bbl@afterfi\csname#2@sh@#1@string##1\endcsname
1962     \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.

```

1963 \def\initiate@active@char#1{%
1964   \bbl@ifunset{active@char\string#1}%
1965   {\bbl@withactive
1966     {\expandafter\@initiate@active@char\expandafter}#1\string#1}%
1967   {}}

```

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).

```

1968 \def\@initiate@active@char#1#2#3{%
1969   \bbl@csarg\edef{oricat@#2}{\catcode`#2=\the\catcode`#2\relax}%
1970   \ifx#1\@undefined
1971     \bbl@csarg\edef{oridef@#2}{\let\noexpand#1\noexpand\@undefined}%
1972   \else
1973     \bbl@csarg\let{oridef@#2}#1%
1974     \bbl@csarg\edef{oridef@#2}{%
1975       \let\noexpand#1%
1976       \expandafter\noexpand\csname bbl@oridef@@#2\endcsname}%
1977   \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*").

```

1978   \ifx#1#3\relax
1979     \expandafter\let\csname normal@char#2\endcsname#3%
1980   \else
1981     \bbl@info{Making #2 an active character}%
1982     \ifnum\mathcode`#2=\ifodd\bbl@engine"1000000 \else"8000 \fi
1983     \@namedef{normal@char#2}{%
1984       \textormath{#3}{\csname bbl@oridef@@#2\endcsname}}%
1985   \else
1986     \@namedef{normal@char#2}{#3}%
1987   \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).

```

1988   \bbl@restoreactive{#2}%
1989   \AtBeginDocument{%
1990     \catcode`#2\active
1991     \if@filesw
1992       \immediate\write\@mainaux{\catcode`\string#2\active}%
1993     \fi}%
1994   \expandafter\bbl@add@special\csname#2\endcsname
1995   \catcode`#2\active
1996   \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⟩`).

```

1997 \let\bbl@tempa\@firstoftwo
1998 \if\string^#2%
1999 \def\bbl@tempa{\noexpand\textormath}%
2000 \else
2001 \ifx\bbl@mathnormal\@undefined\else
2002 \let\bbl@tempa\bbl@mathnormal
2003 \fi
2004 \fi
2005 \expandafter\edef\csname active@char#2\endcsname{%
2006 \bbl@tempa
2007 {\noexpand\if@safe@actives
2008 \noexpand\expandafter
2009 \expandafter\noexpand\csname normal@char#2\endcsname
2010 \noexpand\else
2011 \noexpand\expandafter
2012 \expandafter\noexpand\csname bbl@doactive#2\endcsname
2013 \noexpand\fi}%
2014 {\expandafter\noexpand\csname normal@char#2\endcsname}}%
2015 \bbl@csarg\edef{doactive#2}{%
2016 \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!).

```

2017 \bbl@csarg\edef{active@#2}{%
2018 \noexpand\active@prefix\noexpand#1%
2019 \expandafter\noexpand\csname active@char#2\endcsname}%
2020 \bbl@csarg\edef{normal@#2}{%
2021 \noexpand\active@prefix\noexpand#1%
2022 \expandafter\noexpand\csname normal@char#2\endcsname}%
2023 \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.

```

2024 \bbl@active@def#2\user@group{user@active}{language@active}%
2025 \bbl@active@def#2\language@group{language@active}{system@active}%
2026 \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.

```

2027 \expandafter\edef\csname\user@group @sh@#2@@\endcsname
2028 {\expandafter\noexpand\csname normal@char#2\endcsname}%
2029 \expandafter\edef\csname\user@group @sh@#2@\string\protect@\endcsname
2030 {\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.

```
2031 \if\string'#2%
2032 \let\prim@s\bbl@prim@s
2033 \let\active@math@prime#1%
2034 \fi
2035 \bbl@usehooks{initiateactive}{\#1}{\#2}{\#3}}
```

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

```
2036 <<*More package options>> ≡
2037 \DeclareOption{math=active}{}
2038 \DeclareOption{math=normal}{\def\bbl@mathnormal{\noexpand\textormath}}
2039 <</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*.

```
2040 \@ifpackagewith{babel}{KeepShorthandsActive}%
2041 {\let\bbl@restoreactive\@gobble}%
2042 {\def\bbl@restoreactive#1{%
2043 \bbl@exp{%
2044 \\\AfterBabelLanguage\\CurrentOption
2045 {\catcode`#1=\the\catcode`#1\relax}%
2046 \\\AtEndOfPackage
2047 {\catcode`#1=\the\catcode`#1\relax}}}%
2048 \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.

```
2049 \def\bbl@sh@select#1#2{%
2050 \expandafter\ifx\csname#1@sh@#2@sel\endcsname\relax
2051 \bbl@afterelse\bbl@scndcs
2052 \else
2053 \bbl@afterfi\csname#1@sh@#2@sel\endcsname
2054 \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.

```
2055 \begingroup
2056 \bbl@ifunset{ifincsname}% TODO. Ugly. Correct?
2057 {\gdef\active@prefix#1{%
2058 \ifx\protect\@typeset@protect
2059 \else
2060 \ifx\protect\@unexpandable@protect
2061 \noexpand#1%
2062 \else
2063 \protect#1%
2064 \fi
```

```

2065     \expandafter\@gobble
2066     \fi}}
2067   {\gdef\active@prefix#1{%
2068     \ifincsname
2069       \string#1%
2070       \expandafter\@gobble
2071     \else
2072       \ifx\protect\@typeset@protect
2073       \else
2074         \ifx\protect\@unexpandable@protect
2075           \noexpand#1%
2076         \else
2077           \protect#1%
2078         \fi
2079       \expandafter\expandafter\expandafter\@gobble
2080     \fi
2081   \fi}}
2082 \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⟩`.

```

2083 \newif\if@safe@actives
2084 \@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.

```

2085 \def\bbl@restore@actives{\if@safe@actives\@safe@activesfalse\fi}

```

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

```

2086 \def\bbl@activate#1{%
2087   \bbl@withactive{\expandafter\let\expandafter}#1%
2088   \csname bbl@active@\string#1\endcsname}
2089 \def\bbl@deactivate#1{%
2090   \bbl@withactive{\expandafter\let\expandafter}#1%
2091   \csname bbl@normal@\string#1\endcsname}

```

`\bbl@firstcs` These macros are used only as a trick when declaring shorthands.

```

\bbl@scndcs 2092 \def\bbl@firstcs#1#2{\csname#1\endcsname}
2093 \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.

```

2094 \def\babel@texpdf#1#2#3#4{%
2095   \ifx\texorpdfstring\undefined
2096     \textormath{#1}{#2}%
2097   \else
2098     \texorpdfstring{\textormath{#1}{#3}}{#2}%
2099     % \texorpdfstring{\textormath{#1}{#3}}{\textormath{#2}{#4}}%
2100   \fi}
2101 %
2102 \def\declare@shorthand#1#2{\@decl@short{#1}#2\@nil}
2103 \def\@decl@short#1#2#3\@nil#4{%
2104   \def\bbl@tempa{#3}%
2105   \ifx\bbl@tempa\@empty
2106     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@scndcs
2107     \bbl@ifunset{#1@sh@\string#2@}{}%
2108     {\def\bbl@tempa{#4}%
2109      \expandafter\ifx\csname#1@sh@\string#2@\endcsname\bbl@tempa
2110      \else
2111        \bbl@info
2112          {Redefining #1 shorthand \string#2\%
2113           in language \CurrentOption}%
2114      \fi}%
2115     \@namedef{#1@sh@\string#2@}{#4}%
2116   \else
2117     \expandafter\let\csname #1@sh@\string#2@sel\endcsname\bbl@firstcs
2118     \bbl@ifunset{#1@sh@\string#2@\string#3@}{}%
2119     {\def\bbl@tempa{#4}%
2120      \expandafter\ifx\csname#1@sh@\string#2@\string#3@\endcsname\bbl@tempa
2121      \else
2122        \bbl@info
2123          {Redefining #1 shorthand \string#2\string#3\%
2124           in language \CurrentOption}%
2125      \fi}%
2126     \@namedef{#1@sh@\string#2@\string#3@}{#4}%
2127   \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.

```

2128 \def\textormath{%
2129   \ifmmode
2130     \expandafter\@secondoftwo
2131   \else
2132     \expandafter\@firstoftwo
2133   \fi}

```

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

```

2134 \def\user@group{user}
2135 \def\language@group{english} % TODO. I don't like defaults
2136 \def\system@group{system}

```

`\usesshorthands` 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.

```

2137 \def\usesshorthands{%

```

```

2138 \@ifstar\bb1@usesh@s{\bb1@usesh@x{}}
2139 \def\bb1@usesh@s#1{%
2140 \bb1@usesh@x
2141 {\AddBabelHook{babel-sh-\string#1}{afterextras}{\bb1@activate{#1}}}%
2142 {#1}}
2143 \def\bb1@usesh@x#1#2{%
2144 \bb1@ifshorthand{#2}%
2145 {\def\user@group{user}%
2146 \initiate@active@char{#2}%
2147 #1%
2148 \bb1@activate{#2}}%
2149 {\bb1@error
2150 {Cannot declare a shorthand turned off (\string#2)}
2151 {Sorry, but you cannot use shorthands which have been\\%
2152 turned off in the package options}}}

```

`\defineshorthand` Currently we only support two groups of user level shorthands, named internally `user` and `user<lang>` (language-dependent user shorthands). By default, only the first one is taken into account, but if the former is also used (in the optional argument of `\defineshorthand`) a new level is inserted for it (`user@generic`, done by `\bb1@set@user@generic`); we make also sure `{}` and `\protect` are taken into account in this new top level.

```

2153 \def\user@language@group{user@\language@group}
2154 \def\bb1@set@user@generic#1#2{%
2155 \bb1@ifunset{user@generic@active#1}%
2156 {\bb1@active@def#1\user@language@group{user@active}{user@generic@active}%
2157 \bb1@active@def#1\user@group{user@generic@active}{language@active}%
2158 \expandafter\edef\csname#2@sh#1@@\endcsname{%
2159 \expandafter\noexpand\csname normal@char#1\endcsname}%
2160 \expandafter\edef\csname#2@sh#1@\string\protect\endcsname{%
2161 \expandafter\noexpand\csname user@active#1\endcsname}}%
2162 \@empty}
2163 \newcommand\defineshorthand[3][user]{%
2164 \edef\bb1@tempa{\zap@space#1 \@empty}%
2165 \bb1@for\bb1@tempb\bb1@tempa{%
2166 \if*\expandafter\@car\bb1@tempb\@nil
2167 \edef\bb1@tempb{user\expandafter\@gobble\bb1@tempb}%
2168 \@expandtwoargs
2169 \bb1@set@user@generic{\expandafter\string\@car#2\@nil}\bb1@tempb
2170 \fi
2171 \declare@shorthand{\bb1@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].

```

2172 \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`.

```

2173 \def\aliasshorthand#1#2{%
2174 \bb1@ifshorthand{#2}%
2175 {\expandafter\ifx\csname active@char\string#2\endcsname\relax
2176 \ifx\document\@notprerr
2177 \@notshorthand{#2}%
2178 \else
2179 \initiate@active@char{#2}%

```

```

2180      \expandafter\let\csname active@char\string#2\expandafter\endcsname
2181      \csname active@char\string#1\endcsname
2182      \expandafter\let\csname normal@char\string#2\expandafter\endcsname
2183      \csname normal@char\string#1\endcsname
2184      \bbl@activate{#2}%
2185      \fi
2186    \fi}%
2187  {\bbl@error
2188    {Cannot declare a shorthand turned off (\string#2)}
2189    {Sorry, but you cannot use shorthands which have been\\%
2190     turned off in the package options}}}

```

\@notshorthand

```

2191 \def\@notshorthand#1{%
2192   \bbl@error{%
2193     The character '\string #1' should be made a shorthand character;\\%
2194     add the command \string\usesshorthands\string{#1\string} to
2195     the preamble.\\%
2196     I will ignore your instruction}%
2197   {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.

```

2198 \newcommand*\shorthandon[1]{\bbl@switch@sh\@ne#1\@nnil}
2199 \DeclareRobustCommand*\shorthandoff{%
2200   \@ifstar{\bbl@shorthandoff\tw@}{\bbl@shorthandoff\z@}}
2201 \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.

```

2202 \def\bbl@switch@sh#1#2{%
2203   \ifx#2\@nnil\else
2204     \bbl@ifunset{\bbl@active@\string#2}%
2205     {\bbl@error
2206       {I cannot switch '\string#2' on or off--not a shorthand}%
2207       {This character is not a shorthand. Maybe you made\\%
2208        a typing mistake? I will ignore your instruction}}}%
2209     {\ifcase#1%
2210       \catcode`#2\relax
2211       \or
2212       \catcode`#2\active
2213       \or
2214       \csname bbl@oricat@\string#2\endcsname
2215       \csname bbl@oridef@\string#2\endcsname
2216       \fi}%
2217     \bbl@afterfi\bbl@switch@sh#1%
2218   \fi}

```

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

```

2219 \def\babelshorthand{\active@prefix\babelshorthand\bb@putsh}
2220 \def\bb@putsh#1{%
2221   \bb@ifunset{bb@active@\string#1}%
2222   {\bb@putsh@i#1\@empty\@nnil}%
2223   {\csname bb@active@\string#1\endcsname}}
2224 \def\bb@putsh@i#1#2\@nnil{%
2225   \csname\language@group @sh@\string#1@%
2226   \ifx\@empty#2\else\string#2\fi\endcsname}
2227 \ifx\bb@opt@shorthands\@nnil\else
2228   \let\bb@s@initiate@active@char\initiate@active@char
2229   \def\initiate@active@char#1{%
2230     \bb@ifshorthand{#1}{\bb@s@initiate@active@char{#1}}{}}
2231   \let\bb@s@switch@sh\bb@switch@sh
2232   \def\bb@switch@sh#1#2{%
2233     \ifx#2\@nnil\else
2234       \bb@afterfi
2235       \bb@ifshorthand{#2}{\bb@s@switch@sh#1{#2}}{\bb@switch@sh#1}%
2236     \fi}
2237   \let\bb@s@activate\bb@activate
2238   \def\bb@activate#1{%
2239     \bb@ifshorthand{#1}{\bb@s@activate{#1}}{}}
2240   \let\bb@s@deactivate\bb@deactivate
2241   \def\bb@deactivate#1{%
2242     \bb@ifshorthand{#1}{\bb@s@deactivate{#1}}{}}
2243 \fi

```

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

```

2244 \newcommand\ifbabelshorthand[3]{\bb@ifunset{bb@active@\string#1}{#3}{#2}}

```

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

```

2245 \def\bb@prim@s{%
2246   \prime\futurelet\@let@token\bb@pr@m@s}
2247 \def\bb@if@primes#1#2{%
2248   \ifx#1\@let@token
2249     \expandafter\@firstoftwo
2250   \else\ifx#2\@let@token
2251     \bb@afterelse\expandafter\@firstoftwo
2252   \else
2253     \bb@afterfi\expandafter\@secondoftwo
2254   \fi\fi}
2255 \begingroup
2256   \catcode`\^=7 \catcode`\*=\active \lccode`\*=\^
2257   \catcode`\'=12 \catcode`\"=\active \lccode`\"='\^
2258   \lowercase{%
2259     \gdef\bb@pr@m@s{%
2260       \bb@if@primes"%
2261       \pr@@@s
2262       {\bb@if@primes*\^{\pr@@@t\egroup}}}}
2263 \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).

```
2264 \initiate@active@char{~}
2265 \declare@shorthand{system}{~}{\leavevmode\nobreak\ }
2266 \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.

```
2267 \expandafter\def\csname OT1dpos\endcsname{127}
2268 \expandafter\def\csname T1dpos\endcsname{4}
```

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

```
2269 \ifx\f@encoding\undefined
2270 \def\f@encoding{OT1}
2271 \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.

```
2272 \bbl@trace{Language attributes}
2273 \newcommand\languageattribute[2]{%
2274 \def\bbl@tempc{#1}%
2275 \bbl@fixname\bbl@tempc
2276 \bbl@iflanguage\bbl@tempc{%
2277 \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.

```
2278 \ifx\bbl@known@attribs\undefined
2279 \in@false
2280 \else
2281 \bbl@xin@{,\bbl@tempc-##1,}{,\bbl@known@attribs,}%
2282 \fi
2283 \ifin@
2284 \bbl@warning{%
2285 You have more than once selected the attribute '##1'\%
2286 for language #1. Reported}%
2287 \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.

```
2288 \bbl@exp{%
2289 \\\bbl@add@list\\bbl@known@attribs{\bbl@tempc-##1}}%
2290 \edef\bbl@tempa{\bbl@tempc-##1}%
2291 \expandafter\bbl@ifknown@trib\expandafter{\bbl@tempa}\bbl@attributes%
2292 {\csname\bbl@tempc @attr##1\endcsname}%
2293 {\@attrerr{\bbl@tempc}{##1}}%
2294 \fi}}
2295 \@onlypreamble\languageattribute
```


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

```
2296 \newcommand*{\@attrerr}[2]{%
2297   \bbl@error
2298   {The attribute #2 is unknown for language #1.}%
2299   {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}`.

```
2300 \def\bbl@declare@ttribute#1#2#3{%
2301   \bbl@xin@{,#2,},{,\BabelModifiers,}%
2302   \ifin@
2303     \AfterBabelLanguage{#1}{\languageattribute{#1}{#2}}%
2304   \fi
2305   \bbl@add@list\bbl@attributes{#1-#2}%
2306   \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`'.

```
2307 \def\bbl@ifattributeset#1#2#3#4{%
2308   \ifx\bbl@known@attribs\undefined
2309     \in@false
2310   \else
2311     \bbl@xin@{,#1-#2,},{,\bbl@known@attribs,}%
2312   \fi
2313   \ifin@
2314     \bbl@afterelse#3%
2315   \else
2316     \bbl@afterfi#4%
2317   \fi
2318 }
```

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

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`.

```
2319 \def\bbl@ifknown@ttrib#1#2{%
2320   \let\bbl@tempa\@secondoftwo
2321   \bbl@loopx\bbl@tempb{#2}{%
2322     \expandafter\in@\expandafter{\expandafter,\bbl@tempb,}{,#1,}%
2323   \ifin@
2324     \let\bbl@tempa\@firstoftwo
2325   \else
```

```

2326 \fi}%
2327 \bbl@tempa
2328 }

```

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

```

2329 \def\bbl@clear@ttribs{%
2330 \ifx\bbl@attributes\undefined\else
2331 \bbl@loopx\bbl@tempa{\bbl@attributes}{%
2332 \expandafter\bbl@clear@ttrib\bbl@tempa.
2333 }%
2334 \let\bbl@attributes\undefined
2335 \fi}
2336 \def\bbl@clear@ttrib#1-#2.{%
2337 \expandafter\let\csname#1@attr@#2\endcsname\undefined}
2338 \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`

```

2339 \bbl@trace{Macros for saving definitions}
2340 \def\babel@beginsave{\babel@savecnt\z@}

```

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

```

2341 \newcount\babel@savecnt
2342 \babel@beginsave

```

`\babel@save` The macro `\babel@save<csname>` saves the current meaning of the control sequence `<csname>` to `\originalTeX`³¹. To do this, we let the current meaning to a temporary control sequence, the restore commands are appended to `\originalTeX` and the counter is incremented. The macro `\babel@savevariable<variable>` saves the value of the variable. `<variable>` can be anything allowed after the `\the` primitive.

`\babel@savevariable`

```

2343 \def\babel@save#1{%
2344 \expandafter\let\csname babel@number\babel@savecnt\endcsname#1\relax
2345 \toks@\expandafter{\originalTeX\let#1=}
2346 \bbl@exp{%
2347 \def\\originalTeX{\the\toks@<\babel@number\babel@savecnt>\relax}}%
2348 \advance\babel@savecnt\@ne}
2349 \def\babel@savevariable#1{%
2350 \toks@\expandafter{\originalTeX #1=}
2351 \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.

```

2352 \def\bbl@frenchspacing{%

```

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

```

2353 \ifnum\the\sfcodes\.\.=\@m
2354 \let\bbl@nonfrenchspacing\relax
2355 \else
2356 \frenchspacing
2357 \let\bbl@nonfrenchspacing\nonfrenchspacing
2358 \fi}
2359 \let\bbl@nonfrenchspacing\nonfrenchspacing
2360 %
2361 \let\bbl@elt\relax
2362 \edef\bbl@fs@chars{%
2363 \bbl@elt{\string.}\@m{3000}\bbl@elt{\string?}\@m{3000}%
2364 \bbl@elt{\string!}\@m{3000}\bbl@elt{\string:}\@m{2000}%
2365 \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.

```

2366 \bbl@trace{Short tags}
2367 \def\babeltags#1{%
2368 \edef\bbl@tempa{\zap@space#1 \@empty}%
2369 \def\bbl@tempb##1=##2\@{#1}%
2370 \edef\bbl@tempc{%
2371 \noexpand\newcommand
2372 \expandafter\noexpand\csname ##1\endcsname{%
2373 \noexpand\protect
2374 \expandafter\noexpand\csname otherlanguage*\endcsname{##2}}
2375 \noexpand\newcommand
2376 \expandafter\noexpand\csname text##1\endcsname{%
2377 \noexpand\foreignlanguage{##2}}
2378 \bbl@tempc}%
2379 \bbl@for\bbl@tempa\bbl@tempa{%
2380 \expandafter\bbl@tempb\bbl@tempa\@{#1}}

```

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.

```

2381 \bbl@trace{Hyphens}
2382 \@onlypreamble\babelhyphenation
2383 \AtEndOfPackage{%
2384 \newcommand\babelhyphenation[2][\@empty]{%
2385 \ifx\bbl@hyphenation@\relax
2386 \let\bbl@hyphenation@\@empty
2387 \fi
2388 \ifx\bbl@hyphlist\@empty\else
2389 \bbl@warning{%
2390 You must not intermingle \string\selectlanguage\space and\%
2391 \string\babelhyphenation\space or some exceptions will not\%
2392 be taken into account. Reported}%
2393 \fi
2394 \ifx\@empty#1%
2395 \protected@edef\bbl@hyphenation@{\bbl@hyphenation@\space#2}%
2396 \else

```

```

2397 \bbl@vforeach{#1}{%
2398 \def\bbl@tempa{##1}%
2399 \bbl@fixname\bbl@tempa
2400 \bbl@iflanguage\bbl@tempa{%
2401 \bbl@csarg\protected@edef{hyphenation@\bbl@tempa}{%
2402 \bbl@ifunset{bbl@hyphenation@\bbl@tempa}%
2403 \@empty
2404 {\csname bbl@hyphenation@\bbl@tempa\endcsname\space}%
2405 #2}}}%
2406 \fi}}

```

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

```

2407 \def\bbl@allowhyphens{\ifvmode\else\nobreak\hskip\z@skip\fi}
2408 \def\bbl@t@one{T1}
2409 \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`.

```

2410 \newcommand\babellnullhyphen{\char\hyphenchar\font}
2411 \def\babelhyphen{\active@prefix\babelhyphen\bbl@hyphen}
2412 \def\bbl@hyphen{%
2413 \ifstar{\bbl@hyphen@i @}{\bbl@hyphen@i \@empty}}
2414 \def\bbl@hyphen@i#1#2{%
2415 \bbl@ifunset{bbl@hy@#1#2\@empty}%
2416 {\csname bbl@#1usehyphen\endcsname{\discretionary{#2}{}{#2}}}%
2417 {\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.

```

2418 \def\bbl@usehyphen#1{%
2419 \leavevmode
2420 \ifdim\lastskip>\z@\mbox{#1}\else\nobreak#1\fi
2421 \nobreak\hskip\z@skip}
2422 \def\bbl@@usehyphen#1{%
2423 \leavevmode\ifdim\lastskip>\z@\mbox{#1}\else#1\fi}

```

The following macro inserts the hyphen char.

```

2424 \def\bbl@hyphenchar{%
2425 \ifnum\hyphenchar\font=\m@ne
2426 \babellnullhyphen
2427 \else
2428 \char\hyphenchar\font
2429 \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.

```

2430 \def\bbl@hy@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}{}}{}}
2431 \def\bbl@hy@@soft{\bbl@usehyphen{\discretionary{\bbl@hyphenchar}}{}}
2432 \def\bbl@hy@hard{\bbl@usehyphen\bbl@hyphenchar}

```

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

```

2433 \def\bbl@hy@@hard{\bbl@usehyphen\bbl@hyphenchar}
2434 \def\bbl@hy@nobreak{\bbl@usehyphen{\mbox{\bbl@hyphenchar}}}}
2435 \def\bbl@hy@nobreak{\mbox{\bbl@hyphenchar}}
2436 \def\bbl@hy@repeat{%
2437   \bbl@usehyphen{%
2438     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}}
2439 \def\bbl@hy@@repeat{%
2440   \bbl@usehyphen{%
2441     \discretionary{\bbl@hyphenchar}{\bbl@hyphenchar}{\bbl@hyphenchar}}}}
2442 \def\bbl@hy@empty{\hskip\z@skip}
2443 \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.

```

2444 \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.

```

2445 \bbl@trace{Multiencoding strings}
2446 \def\bbl@tglobal#1{\global\let#1#1}
2447 \def\bbl@recatcode#1{% TODO. Used only once?
2448   \@tempcnta="7F
2449   \def\bbl@tempa{%
2450     \ifnum\@tempcnta>"FF\else
2451       \catcode\@tempcnta=#1\relax
2452       \advance\@tempcnta\@ne
2453       \expandafter\bbl@tempa
2454     \fi}%
2455   \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).

```

2456 \@ifpackagewith{babel}{nocase}%
2457   {\let\bbl@patchuclc\relax}%
2458   {\def\bbl@patchuclc{%
2459     \global\let\bbl@patchuclc\relax
2460     \g@addto@macro\@uclclist{\reserved@b{\reserved@b\bbl@uclc}}}%
2461     \gdef\bbl@uclc##1{%
2462       \let\bbl@encoded\bbl@encoded@uclc
2463       \bbl@ifunset{\language @bbl@uclc}% and resumes it
2464       {##1}%

```

```

2465      {\let\bbl@tempa##1\relax % Used by LANG@bbl@uc1c
2466      \csname\language @bbl@uc1c\endcsname}%
2467      {\bbl@tolower\@empty}{\bbl@toupper\@empty}}}%
2468      \gdef\bbl@tolower{\csname\language @bbl@lc\endcsname}%
2469      \gdef\bbl@toupper{\csname\language @bbl@uc\endcsname}}}%
2470 <<(*More package options)>> ≡
2471 \DeclareOption{nocase}{}
2472 <</More package options>>

```

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

```

2473 <<(*More package options)>> ≡
2474 \let\bbl@opt@strings\@nnil % accept strings=value
2475 \DeclareOption{strings}{\def\bbl@opt@strings{\BabelStringsDefault}}
2476 \DeclareOption{strings=encoded}{\let\bbl@opt@strings\relax}
2477 \def\BabelStringsDefault{generic}
2478 <</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.

```

2479 \@onlypreamble\StartBabelCommands
2480 \def\StartBabelCommands{%
2481   \begingroup
2482   \bbl@recatcode{11}%
2483   <<Macros local to BabelCommands>>
2484   \def\bbl@provstring##1##2{%
2485     \providecommand##1{##2}%
2486     \bbl@tglobal##1}%
2487   \global\let\bbl@scafter\@empty
2488   \let\StartBabelCommands\bbl@startcmds
2489   \ifx\BabelLanguages\relax
2490     \let\BabelLanguages\CurrentOption
2491   \fi
2492   \begingroup
2493   \let\bbl@screset\@nnil % local flag - disable 1st stopcommands
2494   \StartBabelCommands}
2495 \def\bbl@startcmds{%
2496   \ifx\bbl@screset\@nnil\else
2497     \bbl@usehooks{stopcommands}{}%
2498   \fi
2499   \endgroup
2500   \begingroup
2501   \@ifstar
2502     {\ifx\bbl@opt@strings\@nnil
2503       \let\bbl@opt@strings\BabelStringsDefault
2504     \fi
2505     \bbl@startcmds@i}%
2506   \bbl@startcmds@i}
2507 \def\bbl@startcmds@i#1#2{%
2508   \edef\bbl@L{\zap@space#1 \@empty}%
2509   \edef\bbl@G{\zap@space#2 \@empty}%
2510   \bbl@startcmds@ii}
2511 \let\bbl@startcmds\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.

```

2512 \newcommand\bbbl@startcmds@ii[1][\@empty]{%
2513   \let\SetString\@gobbletwo
2514   \let\bbbl@stringdef\@gobbletwo
2515   \let\AfterBabelCommands\@gobble
2516   \ifx\@empty#1%
2517     \def\bbbl@sc@label{generic}%
2518     \def\bbbl@encstring##1##2{%
2519       \ProvideTextCommandDefault##1{##2}%
2520       \bbbl@tglobal##1%
2521       \expandafter\bbbl@tglobal\csname\string?\string##1\endcsname}%
2522     \let\bbbl@sctest\in@true
2523   \else
2524     \let\bbbl@sc@charset\space % <- zapped below
2525     \let\bbbl@sc@fontenc\space % <- " "
2526     \def\bbbl@tempa##1=##2\@nil{%
2527       \bbbl@csarg\edef{sc@\zap@space##1 \@empty}{##2 }}%
2528     \bbbl@foreach{label=#1}{\bbbl@tempa##1\@nil}%
2529     \def\bbbl@tempa##1 ##2{% space -> comma
2530       ##1%
2531       \ifx\@empty##2\else\ifx,##1,\else,\fi\bbbl@afterfi\bbbl@tempa##2\fi}%
2532     \edef\bbbl@sc@fontenc{\expandafter\bbbl@tempa\bbbl@sc@fontenc\@empty}%
2533     \edef\bbbl@sc@label{\expandafter\zap@space\bbbl@sc@label\@empty}%
2534     \edef\bbbl@sc@charset{\expandafter\zap@space\bbbl@sc@charset\@empty}%
2535     \def\bbbl@encstring##1##2{%
2536       \bbbl@foreach\bbbl@sc@fontenc{%
2537         \bbbl@ifunset{T@####1}%
2538         {}%
2539         {\ProvideTextCommand##1{####1}{##2}%
2540         \bbbl@tglobal##1%
2541         \expandafter
2542         \bbbl@tglobal\csname####1\string##1\endcsname}}}%
2543     \def\bbbl@sctest{%
2544       \bbbl@xin@{\bbbl@opt@strings,}{,\bbbl@sc@label,\bbbl@sc@fontenc,}}%
2545   \fi
2546   \ifx\bbbl@opt@strings\@nnil % ie, no strings key -> defaults
2547   \else\ifx\bbbl@opt@strings\relax % ie, strings=encoded
2548     \let\AfterBabelCommands\bbbl@aftercmds
2549     \let\SetString\bbbl@setstring
2550     \let\bbbl@stringdef\bbbl@encstring
2551   \else % ie, strings=value
2552     \bbbl@sctest
2553   \fin@
2554     \let\AfterBabelCommands\bbbl@aftercmds
2555     \let\SetString\bbbl@setstring
2556     \let\bbbl@stringdef\bbbl@provstring
2557   \fi\fi\fi
2558   \bbbl@scswitch
2559   \ifx\bbbl@G\@empty
2560     \def\SetString##1##2{%
2561       \bbbl@error{Missing group for string \string##1}%
2562       {You must assign strings to some category, typically\\%
2563       captions or extras, but you set none}}%

```

```

2564 \fi
2565 \ifx\@empty#1%
2566 \bbl@usehooks{defaultcommands}{}%
2567 \else
2568 \@expandtwoargs
2569 \bbl@usehooks{encodedcommands}{\bbl@sc@charset}{\bbl@sc@fontenc}}%
2570 \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) .

```

2571 \def\bbl@forlang#1#2{%
2572 \bbl@for#1\bbl@L{%
2573 \bbl@xin@{,#1,},{,\BabelLanguages,}%
2574 \ifin@#2\relax\fi}}
2575 \def\bbl@scswitch{%
2576 \bbl@forlang\bbl@tempa{%
2577 \ifx\bbl@G\@empty\else
2578 \ifx\SetString\@gobbletwo\else
2579 \edef\bbl@GL{\bbl@G\bbl@tempa}%
2580 \bbl@xin@{,\bbl@GL,},{,\bbl@screset,}%
2581 \ifin@\else
2582 \global\expandafter\let\csname\bbl@GL\endcsname\@undefined
2583 \xdef\bbl@screset{\bbl@screset,\bbl@GL}%
2584 \fi
2585 \fi
2586 \fi}}
2587 \AtEndOfPackage{%
2588 \def\bbl@forlang#1#2{\bbl@for#1\bbl@L{\bbl@ifunset{date#1}{\#2}}}%
2589 \let\bbl@scswitch\relax}
2590 \@onlypreamble\EndBabelCommands
2591 \def\EndBabelCommands{%
2592 \bbl@usehooks{stopcommands}{}%
2593 \endgroup
2594 \endgroup
2595 \bbl@scafter}
2596 \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.

```

2597 \def\bbl@setstring#1#2{% eg, \prefacename{<string>}
2598 \bbl@forlang\bbl@tempa{%
2599 \edef\bbl@LC{\bbl@tempa\bbl@stripslash#1}%
2600 \bbl@ifunset{\bbl@LC}% eg, \germanchaptername
2601 {\bbl@exp{%
2602 \global\bbbl@add\<\bbl@G\bbl@tempa>\bbbl@scset\#1\<\bbl@LC>}}}%

```



```

2603     {}%
2604     \def\BabelString{#2}%
2605     \bbl@usehooks{stringprocess}{}%
2606     \expandafter\bbl@stringdef
2607     \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`.

```

2608 \ifx\bbl@opt@strings\relax
2609   \def\bbl@scset#1#2{\def#1{\bbl@encoded#2}}
2610   \bbl@patchuclc
2611   \let\bbl@encoded\relax
2612   \def\bbl@encoded@uclc#1{%
2613     \@inmathwarn#1%
2614     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
2615       \expandafter\ifx\csname ?\string#1\endcsname\relax
2616         \TextSymbolUnavailable#1%
2617       \else
2618         \csname ?\string#1\endcsname
2619       \fi
2620     \else
2621       \csname\cf@encoding\string#1\endcsname
2622     \fi}
2623 \else
2624   \def\bbl@scset#1#2{\def#1{#2}}
2625 \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.

```

2626 <<*Macros local to BabelCommands>> ≡
2627 \def\SetStringLoop##1##2{%
2628   \def\bbl@templ####1{\expandafter\noexpand\csname##1\endcsname}%
2629   \count@\z@
2630   \bbl@loop\bbl@tempa{##2}{% empty items and spaces are ok
2631     \advance\count@\@ne
2632     \toks@\expandafter{\bbl@tempa}%
2633     \bbl@exp{%
2634       \\SetString\bbl@templ{\romannumeral\count@}{\the\toks@}%
2635       \count@=\the\count@\relax}}}%
2636 <</Macros local to BabelCommands>>

```

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

```

2637 \def\bbl@aftercmds#1{%
2638   \toks@\expandafter{\bbl@scafter#1}%
2639   \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.

```

2640 <<*Macros local to BabelCommands>> ≡
2641 \newcommand\SetCase[3][{}%
2642   \bbl@patchuclc

```

```

2643 \bbl@forlang\bbl@tempa{%
2644 \expandafter\bbl@encstring
2645 \csname\bbl@tempa @bbl@ucl\endcsname{\bbl@tempa##1}%
2646 \expandafter\bbl@encstring
2647 \csname\bbl@tempa @bbl@uc\endcsname{##2}%
2648 \expandafter\bbl@encstring
2649 \csname\bbl@tempa @bbl@lc\endcsname{##3}}}%
2650 <</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.

```

2651 <<*Macros local to BabelCommands>> ≡
2652 \newcommand\SetHyphenMap[1]{%
2653 \bbl@forlang\bbl@tempa{%
2654 \expandafter\bbl@stringdef
2655 \csname\bbl@tempa @bbl@hyphenmap\endcsname{##1}}}%
2656 <</Macros local to BabelCommands>>

```

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

```

2657 \newcommand\BabelLower[2]{% one to one.
2658 \ifnum\lccode#1=#2\else
2659 \babel@savevariable{\lccode#1}%
2660 \lccode#1=#2\relax
2661 \fi}
2662 \newcommand\BabelLowerMM[4]{% many-to-many
2663 \@tempcnta=#1\relax
2664 \@tempcntb=#4\relax
2665 \def\bbl@tempa{%
2666 \ifnum\@tempcnta>#2\else
2667 \@expandtwoargs\BabelLower{\the\@tempcnta}{\the\@tempcntb}%
2668 \advance\@tempcnta#3\relax
2669 \advance\@tempcntb#3\relax
2670 \expandafter\bbl@tempa
2671 \fi}%
2672 \bbl@tempa}
2673 \newcommand\BabelLowerMO[4]{% many-to-one
2674 \@tempcnta=#1\relax
2675 \def\bbl@tempa{%
2676 \ifnum\@tempcnta>#2\else
2677 \@expandtwoargs\BabelLower{\the\@tempcnta}{#4}%
2678 \advance\@tempcnta#3
2679 \expandafter\bbl@tempa
2680 \fi}%
2681 \bbl@tempa}

```

The following package options control the behavior of hyphenation mapping.

```

2682 <<*More package options>> ≡
2683 \DeclareOption{hyphenmap=off}{\chardef\bbl@opt@hyphenmap\z@}
2684 \DeclareOption{hyphenmap=first}{\chardef\bbl@opt@hyphenmap\@ne}
2685 \DeclareOption{hyphenmap=select}{\chardef\bbl@opt@hyphenmap\tw@}
2686 \DeclareOption{hyphenmap=other}{\chardef\bbl@opt@hyphenmap\thr@@}
2687 \DeclareOption{hyphenmap=other*}{\chardef\bbl@opt@hyphenmap4\relax}
2688 <</More package options>>

```

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

```

2689 \AtEndOfPackage{%
2690 \ifx\bbl@opt@hyphenmap\@undefined
2691 \bbl@xin@{,}{\bbl@language@opts}%

```

```

2692 \chardef\bbl@opt@hyphenmap\ifin@4\else\@ne\fi
2693 \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.

```

2694 \newcommand\setlocalecaption{% TODO. Catch typos. What about ensure?
2695 \@ifstar\bbl@setcaption@s\bbl@setcaption@x}
2696 \def\bbl@setcaption@x#1#2#3{% language caption-name string
2697 \edef\bbl@tempa{#1}%
2698 \edef\bbl@tempd{%
2699 \expandafter\expandafter\expandafter
2700 \strip@prefix\expandafter\meaning\csname captions#1\endcsname}%
2701 \bbl@xin@
2702 {\expandafter\string\csname #2name\endcsname}%
2703 {\bbl@tempd}%
2704 \ifin@ % Renew caption
2705 \bbl@xin@{\string\bbl@scset}{\bbl@tempd}%
2706 \ifin@
2707 \bbl@exp{%
2708 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2709 {\\\bbl@scset\<#2name>\<#1#2name>}%
2710 {}}%
2711 \else % Old way converts to new way
2712 \bbl@ifunset{#1#2name}%
2713 {\bbl@exp{%
2714 \\\bbl@add\<captions#1>\def\<#2name>\<#1#2name>}}%
2715 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2716 {\def\<#2name>\<#1#2name>}}%
2717 {}}%
2718 {}%
2719 \fi
2720 \else
2721 \bbl@xin@{\string\bbl@scset}{\bbl@tempd}% New
2722 \ifin@ % New way
2723 \bbl@exp{%
2724 \\\bbl@add\<captions#1>\\\bbl@scset\<#2name>\<#1#2name>}%
2725 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2726 {\\\bbl@scset\<#2name>\<#1#2name>}%
2727 {}}%
2728 \else % Old way, but defined in the new way
2729 \bbl@exp{%
2730 \\\bbl@add\<captions#1>\def\<#2name>\<#1#2name>}}%
2731 \\\bbl@ifsamestring{\bbl@tempa}{\language}%
2732 {\def\<#2name>\<#1#2name>}}%
2733 {}}%
2734 \fi%
2735 \fi
2736 \@namedef{#1#2name}{#3}%
2737 \toks@ \expandafter\bbl@captionslist}%
2738 \bbl@exp{\in@{\<#2name>}{\the\toks@}}%
2739 \ifin@ \else
2740 \bbl@exp{\\\bbl@add\\bbl@captionslist{\<#2name>}}%
2741 \bbl@tglobal\bbl@captionslist
2742 \fi}
2743 % \def\bbl@setcaption@s#1#2#3{} % 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.

```
2744 \bbl@trace{Macros related to glyphs}
2745 \def\set@low@box#1{\setbox\tw@hbox{,}\setbox\z@hbox{#1}%
2746   \dimen\z@ht\z@ \advance\dimen\z@ -ht\tw@%
2747   \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.

```
2748 \def\save@sf@q#1{\leavevmode
2749   \begingroup
2750   \edef\@SF{\spacefactor\the\spacefactor}#1\@SF
2751   \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.

```
2752 \ProvideTextCommand{\quotedblbase}{OT1}{%
2753   \save@sf@q{\set@low@box{\textquotedblright\}}%
2754   \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.

```
2755 \ProvideTextCommandDefault{\quotedblbase}{%
2756   \UseTextSymbol{OT1}{\quotedblbase}}
```

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

```
2757 \ProvideTextCommand{\quotesinglbase}{OT1}{%
2758   \save@sf@q{\set@low@box{\textquoteright\}}%
2759   \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.

```
2760 \ProvideTextCommandDefault{\quotesinglbase}{%
2761   \UseTextSymbol{OT1}{\quotesinglbase}}
```

`\guillemetleft` The guillemet characters are not available in OT1 encoding. They are faked. (Wrong names
`\guillemetright` with o preserved for compatibility.)

```
2762 \ProvideTextCommand{\guillemetleft}{OT1}{%
2763   \ifmmode
2764     \ll
2765   \else
2766     \save@sf@q{\nobreak
2767       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
2768     \fi}
2769 \ProvideTextCommand{\guillemetright}{OT1}{%
2770   \ifmmode
2771     \gg
```

```

2772 \else
2773   \save@sf@q{\nobreak
2774     \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
2775   \fi}
2776 \ProvideTextCommand{\guillemotleft}{OT1}{%
2777   \ifmmode
2778     \ll
2779   \else
2780     \save@sf@q{\nobreak
2781       \raise.2ex\hbox{$\scriptscriptstyle\ll$}\bbl@allowhyphens}%
2782     \fi}
2783 \ProvideTextCommand{\guillemotright}{OT1}{%
2784   \ifmmode
2785     \gg
2786   \else
2787     \save@sf@q{\nobreak
2788       \raise.2ex\hbox{$\scriptscriptstyle\gg$}\bbl@allowhyphens}%
2789     \fi}

```

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

```

2790 \ProvideTextCommandDefault{\guillemetleft}{%
2791   \UseTextSymbol{OT1}{\guillemetleft}}
2792 \ProvideTextCommandDefault{\guillemetright}{%
2793   \UseTextSymbol{OT1}{\guillemetright}}
2794 \ProvideTextCommandDefault{\guillemotleft}{%
2795   \UseTextSymbol{OT1}{\guillemotleft}}
2796 \ProvideTextCommandDefault{\guillemotright}{%
2797   \UseTextSymbol{OT1}{\guillemotright}}

```

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

```

2798 \ProvideTextCommand{\guilsinglleft}{OT1}{%
2799   \ifmmode
2800     <%
2801   \else
2802     \save@sf@q{\nobreak
2803       \raise.2ex\hbox{$\scriptscriptstyle<$}\bbl@allowhyphens}%
2804     \fi}
2805 \ProvideTextCommand{\guilsinglright}{OT1}{%
2806   \ifmmode
2807     >%
2808   \else
2809     \save@sf@q{\nobreak
2810       \raise.2ex\hbox{$\scriptscriptstyle>$}\bbl@allowhyphens}%
2811     \fi}

```

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

```

2812 \ProvideTextCommandDefault{\guilsinglleft}{%
2813   \UseTextSymbol{OT1}{\guilsinglleft}}
2814 \ProvideTextCommandDefault{\guilsinglright}{%
2815   \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
`\IJ` encoded fonts. Therefore we fake it for the OT1 encoding.

```

2816 \DeclareTextCommand{\ij}{OT1}{%

```

```

2817 i\kern-0.02em\bbl@allowhyphens j}
2818 \DeclareTextCommand{\IJ}{OT1}{%
2819 I\kern-0.02em\bbl@allowhyphens J}
2820 \DeclareTextCommand{\ij}{T1}{\char188}
2821 \DeclareTextCommand{\IJ}{T1}{\char156}

```

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

```

2822 \ProvideTextCommandDefault{\ij}{%
2823 \UseTextSymbol{OT1}{\ij}}
2824 \ProvideTextCommandDefault{\IJ}{%
2825 \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).

```

2826 \def\crrtic@{\hrule height0.1ex width0.3em}
2827 \def\crttic@{\hrule height0.1ex width0.33em}
2828 \def\ddj@{%
2829 \setbox0\hbox{d}\dimen@=\ht0
2830 \advance\dimen@1ex
2831 \dimen@.45\dimen@
2832 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2833 \advance\dimen@ii.5ex
2834 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crrtic@}}}}
2835 \def\DDJ@{%
2836 \setbox0\hbox{D}\dimen@=.55\ht0
2837 \dimen@ii\expandafter\rem@pt\the\fontdimen\@ne\font\dimen@
2838 \advance\dimen@ii.15ex % correction for the dash position
2839 \advance\dimen@ii-.15\fontdimen7\font % correction for cmtt font
2840 \dimen\thr@@\expandafter\rem@pt\the\fontdimen7\font\dimen@
2841 \leavevmode\rlap{\raise\dimen@\hbox{\kern\dimen@ii\vbox{\crttic@}}}}
2842 %
2843 \DeclareTextCommand{\dj}{OT1}{\ddj@ d}
2844 \DeclareTextCommand{\DJ}{OT1}{\DDJ@ D}

```

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

```

2845 \ProvideTextCommandDefault{\dj}{%
2846 \UseTextSymbol{OT1}{\dj}}
2847 \ProvideTextCommandDefault{\DJ}{%
2848 \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.

```

2849 \DeclareTextCommand{\SS}{OT1}{SS}
2850 \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 2851 \ProvideTextCommandDefault{\glq}{%
2852 \textormath{\quotesinglbase}{\mbox{\quotesinglbase}}}

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

2853 \ProvideTextCommand{\grq}{T1}{%
2854 \textormath{\kern\z@\textquoteleft}{\mbox{\textquoteleft}}}
2855 \ProvideTextCommand{\grq}{TU}{%
2856 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}
2857 \ProvideTextCommand{\grq}{OT1}{%
2858 \save@sf@q{\kern-.0125em
2859 \textormath{\textquoteleft}{\mbox{\textquoteleft}}}%
2860 \kern.07em\relax}}
2861 \ProvideTextCommandDefault{\grq}{\UseTextSymbol{OT1}\grq}
```

`\glqq` The ‘german’ double quotes.

```
\grqq 2862 \ProvideTextCommandDefault{\glqq}{%
2863 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}

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

2864 \ProvideTextCommand{\grqq}{T1}{%
2865 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2866 \ProvideTextCommand{\grqq}{TU}{%
2867 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
2868 \ProvideTextCommand{\grqq}{OT1}{%
2869 \save@sf@q{\kern-.07em
2870 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}%
2871 \kern.07em\relax}}
2872 \ProvideTextCommandDefault{\grqq}{\UseTextSymbol{OT1}\grqq}
```

`\flq` The ‘french’ single guillemets.

```
\frq 2873 \ProvideTextCommandDefault{\flq}{%
2874 \textormath{\guilsinglleft}{\mbox{\guilsinglleft}}}
2875 \ProvideTextCommandDefault{\frq}{%
2876 \textormath{\guilsinglright}{\mbox{\guilsinglright}}}
```

`\flqq` The ‘french’ double guillemets.

```
\frqq 2877 \ProvideTextCommandDefault{\flqq}{%
2878 \textormath{\guillemetleft}{\mbox{\guillemetleft}}}
2879 \ProvideTextCommandDefault{\frqq}{%
2880 \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
`\umlautlow` positioning, the default will be `\umlauthigh` (the normal positioning).

```
2881 \def\umlauthigh{%
2882 \def\bb1@umlauta##1{\leavevmode\bgroup%
2883 \expandafter\accent\csname\f@encoding dqpos\endcsname
```

```

2884      ##1\bb1@allowhyphens\egroup}%
2885 \let\bb1@umlaut\bb1@umlauta}
2886 \def\umlautlow{%
2887 \def\bb1@umlauta{\protect\lower@umlaut}}
2888 \def\umlautelow{%
2889 \def\bb1@umlaute{\protect\lower@umlaut}}
2890 \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.

```

2891 \expandafter\ifx\csname U@D\endcsname\relax
2892 \csname newdimen\endcsname\U@D
2893 \fi

```

The following code fools T_EX's make_accent procedure about the current x-height of the font to force another placement of the umlaut character. First we have to save the current x-height of the font, because we'll change this font dimension and this is always done globally.

Then we compute the new x-height in such a way that the umlaut character is lowered to the base character. The value of .45ex depends on the METAFONT parameters with which the fonts were built. (Just try out, which value will look best.) If the new x-height is too low, it is not changed. Finally we call the \accent primitive, reset the old x-height and insert the base character in the argument.

```

2894 \def\lower@umlaut#1{%
2895 \leavevmode\bgroup
2896 \U@D 1ex%
2897 {\setbox\z@\hbox{%
2898 \expandafter\char\csname\fontencoding dqpos\endcsname}%
2899 \dimen@ -.45ex\advance\dimen@\ht\z@
2900 \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
2901 \expandafter\accent\csname\fontencoding dqpos\endcsname
2902 \fontdimen5\font\U@D #1%
2903 \egroup}

```

For all vowels we declare \" to be a composite command which uses \bb1@umlauta or \bb1@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 \bb1@umlauta and/or \bb1@umlaute for a language in the corresponding ldf (using the babel switching mechanism, of course).

```

2904 \AtBeginDocument{%
2905 \DeclareTextCompositeCommand{\"}{OT1}{a}{\bb1@umlauta{a}}%
2906 \DeclareTextCompositeCommand{\"}{OT1}{e}{\bb1@umlaute{e}}%
2907 \DeclareTextCompositeCommand{\"}{OT1}{i}{\bb1@umlaute{i}}%
2908 \DeclareTextCompositeCommand{\"}{OT1}{i}{\bb1@umlaute{i}}%
2909 \DeclareTextCompositeCommand{\"}{OT1}{o}{\bb1@umlauta{o}}%
2910 \DeclareTextCompositeCommand{\"}{OT1}{u}{\bb1@umlauta{u}}%
2911 \DeclareTextCompositeCommand{\"}{OT1}{A}{\bb1@umlauta{A}}%
2912 \DeclareTextCompositeCommand{\"}{OT1}{E}{\bb1@umlaute{E}}%
2913 \DeclareTextCompositeCommand{\"}{OT1}{I}{\bb1@umlaute{I}}%
2914 \DeclareTextCompositeCommand{\"}{OT1}{O}{\bb1@umlauta{O}}%
2915 \DeclareTextCompositeCommand{\"}{OT1}{U}{\bb1@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.


```

2916 \ifx\l@english\@undefined
2917   \chardef\l@english\z@
2918 \fi
2919 % The following is used to cancel rules in ini files (see Amharic).
2920 \ifx\l@babelnohyphens\@undefined
2921   \newlanguage\l@babelnohyphens
2922 \fi

```

9.13 Layout

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

```

2923 \bbl@trace{Bidi layout}
2924 \providecommand\IfBabelLayout[3]{#3}%
2925 \newcommand\BabelPatchSection[1]{%
2926   \@ifundefined{#1}{}{%
2927     \bbl@exp{\let\<bbl@ss@#1>\<#1>}%
2928     \@namedef{#1}{%
2929       \ifstar{\bbl@presec@#1}%
2930       {\@dblarg{\bbl@presec@x{#1}}}}%
2931 \def\bbl@presec@x#1[#2]#3{%
2932   \bbl@exp{%
2933     \\\select@language@x{\bbl@main@language}%
2934     \\\bbl@cs{sspre@#1}%
2935     \\\bbl@cs{ss@#1}%
2936     [\\foreignlanguage{\language}{\unexpanded{#2}}}%
2937     {\\foreignlanguage{\language}{\unexpanded{#3}}}%
2938     \\\select@language@x{\language}}}%
2939 \def\bbl@presec@#1#2{%
2940   \bbl@exp{%
2941     \\\select@language@x{\bbl@main@language}%
2942     \\\bbl@cs{sspre@#1}%
2943     \\\bbl@cs{ss@#1}*%
2944     {\\foreignlanguage{\language}{\unexpanded{#2}}}%
2945     \\\select@language@x{\language}}}%
2946 \IfBabelLayout{sectioning}%
2947   {\BabelPatchSection{part}%
2948    \BabelPatchSection{chapter}%
2949    \BabelPatchSection{section}%
2950    \BabelPatchSection{subsection}%
2951    \BabelPatchSection{subsubsection}%
2952    \BabelPatchSection{paragraph}%
2953    \BabelPatchSection{subparagraph}%
2954    \def\babel@toc#1{%
2955      \select@language@x{\bbl@main@language}}}%
2956 \IfBabelLayout{captions}%
2957   {\BabelPatchSection{caption}}%

```

9.14 Load engine specific macros

```

2958 \bbl@trace{Input engine specific macros}
2959 \ifcase\bbl@engine
2960   \input txtbabel.def
2961 \or
2962   \input luababel.def
2963 \or
2964   \input xebabel.def
2965 \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.

```
2966 \bbl@trace{Creating languages and reading ini files}
2967 \newcommand\babelprovide[2][{}]{%
2968   \let\bbl@savelangname\language
2969   \edef\bbl@savelocaleid{\the\localeid}%
2970   % Set name and locale id
2971   \edef\language{#2}%
2972   % \global\@namedef\bbl@lcname{#2}{#2}%
2973   \bbl@id@assign
2974   \let\bbl@KVP@captions\@nil
2975   \let\bbl@KVP@date\@nil
2976   \let\bbl@KVP@import\@nil
2977   \let\bbl@KVP@main\@nil
2978   \let\bbl@KVP@script\@nil
2979   \let\bbl@KVP@language\@nil
2980   \let\bbl@KVP@hyphenrules\@nil
2981   \let\bbl@KVP@mapfont\@nil
2982   \let\bbl@KVP@maparabic\@nil
2983   \let\bbl@KVP@mapdigits\@nil
2984   \let\bbl@KVP@intraspace\@nil
2985   \let\bbl@KVP@intrapenalty\@nil
2986   \let\bbl@KVP@onchar\@nil
2987   \let\bbl@KVP@alph\@nil
2988   \let\bbl@KVP@Alph\@nil
2989   \let\bbl@KVP@labels\@nil
2990   \bbl@csarg\let{KVP@labels*}\@nil
2991   \bbl@forkv{#1}{% TODO - error handling
2992     \in@{/{}}{##1}%
2993     \ifin@
2994       \bbl@renewinikey##1\@{##2}%
2995     \else
2996       \bbl@csarg\def{KVP@##1}{##2}%
2997     \fi}%
2998   % == import, captions ==
2999   \ifx\bbl@KVP@import\@nil\else
3000     \bbl@exp{\bbl@ifblank{\bbl@KVP@import}}%
3001     {\ifx\bbl@initoload\relax
3002       \begingroup
3003         \def\BabelBeforeIni##1##2{\gdef\bbl@KVP@import{##1}\endinput}%
3004         \bbl@input@texini{##2}%
3005       \endgroup
3006     \else
3007       \xdef\bbl@KVP@import{\bbl@initoload}%
3008     \fi}%
3009   {}%
3010 \fi
3011 \ifx\bbl@KVP@captions\@nil
3012   \let\bbl@KVP@captions\bbl@KVP@import
3013 \fi
3014 % Load ini
3015 \bbl@ifunset{date#2}%
3016   {\bbl@provide@new{#2}}%
3017   {\bbl@ifblank{#1}%
3018     {\bbl@error
```

```

3019         {If you want to modify `#2' you must tell how in\\%
3020         the optional argument. See the manual for the\\%
3021         available options.}%
3022         {Use this macro as documented}}%
3023         {\bbl@provide@renew{#2}}}%
3024 % Post tasks
3025 \bbl@ifunset{\bbl@extracaps@#2}%
3026     {\bbl@exp{\babelensure[exclude=\\today]{#2}}}%
3027     {\toks@\\expandafter\\expandafter\\expandafter
3028     {\csname bbl@extracaps@#2\endcsname}%
3029     \bbl@exp{\babelensure[exclude=\\today,include=\\the\toks@]{#2}}}%
3030 \bbl@ifunset{\bbl@ensure@\\language}%
3031     {\bbl@exp{%
3032         \\DeclareRobustCommand<\bbl@ensure@\\language>[1]{%
3033         \\foreignlanguage{\\language}%
3034         {###1}}}%
3035     }%
3036 \bbl@exp{%
3037     \\bbl@tglobal\\<\bbl@ensure@\\language>%
3038     \\bbl@tglobal\\<\bbl@ensure@\\language\\space>%
3039 % At this point all parameters are defined if 'import'. Now we
3040 % execute some code depending on them. But what about if nothing was
3041 % imported? We just load the very basic parameters.
3042 \bbl@load@basic{#2}%
3043 % == script, language ==
3044 % Override the values from ini or defines them
3045 \ifx\bbl@KVP@script\\nil\\else
3046     \bbl@csarg\\edef{sname@#2}{\bbl@KVP@script}%
3047 \fi
3048 \ifx\bbl@KVP@language\\nil\\else
3049     \bbl@csarg\\edef{lname@#2}{\bbl@KVP@language}%
3050 \fi
3051 % == onchar ==
3052 \ifx\bbl@KVP@onchar\\nil\\else
3053     \bbl@luahyphenate
3054     \directlua{
3055         if Babel.locale_mapped == nil then
3056             Babel.locale_mapped = true
3057             Babel.linebreaking.add_before(Babel.locale_map)
3058             Babel.loc_to_scr = {}
3059             Babel.chr_to_loc = Babel.chr_to_loc or {}
3060         end}%
3061 \bbl@xin@{ ids }{ \bbl@KVP@onchar\\space}%
3062 \ifin@
3063     \ifx\bbl@starthyphens\\undefined % Needed if no explicit selection
3064         \AddBabelHook{babel-onchar}{beforestart}{\bbl@starthyphens}%
3065     \fi
3066     \bbl@exp{\bbl@add\bbl@starthyphens
3067     {\bbl@patterns@lua{\\language}}}%
3068 % TODO - error/warning if no script
3069     \directlua{
3070         if Babel.script_blocks['\bbl@cl{sbcp}'] then
3071             Babel.loc_to_scr[\\the\\localeid] =
3072             Babel.script_blocks['\bbl@cl{sbcp}']
3073             Babel.locale_props[\\the\\localeid].lc = \\the\\localeid\\space
3074             Babel.locale_props[\\the\\localeid].lg = \\the\\nameuse{l@\\language}\\space
3075         end
3076     }%
3077 \fi

```

```

3078 \bbl@xin@{ fonts }{ \bbl@KVP@onchar\space}%
3079 \ifin@
3080 \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
3081 \bbl@ifunset{\bbl@wdir@\language}\bbl@provide@dirs{\language}}{}%
3082 \directlua{
3083   if Babel.script_blocks['\bbl@cl{sbc}'] then
3084     Babel.loc_to_scr[\the\localeid] =
3085       Babel.script_blocks['\bbl@cl{sbc}']
3086   end}%
3087 \ifx\bbl@mapselect\undefined
3088   \AtBeginDocument{%
3089     \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}}%
3090     {\selectfont}}%
3091   \def\bbl@mapselect{%
3092     \let\bbl@mapselect\relax
3093     \edef\bbl@prefontid{\fontid\font}}%
3094   \def\bbl@mapdir##1{%
3095     {\def\language{##1}%
3096       \let\bbl@ifrestoring\@firstoftwo % To avoid font warning
3097       \bbl@switchfont
3098       \directlua{
3099         Babel.locale_props[\the\csname bbl@id@##1\endcsname]
3100           [\bbl@prefontid] = \fontid\font\space}}}%
3101   \fi
3102   \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
3103   \fi
3104   % TODO - catch non-valid values
3105 \fi
3106 % == mapfont ==
3107 % For bidi texts, to switch the font based on direction
3108 \ifx\bbl@KVP@mapfont\@nil\else
3109   \bbl@ifsamestring{\bbl@KVP@mapfont}{direction}}{}%
3110   {\bbl@error{Option '\bbl@KVP@mapfont' unknown for\the
3111     mapfont. Use 'direction'.%
3112     {See the manual for details.}}}%
3113   \bbl@ifunset{\bbl@lsys@\language}\bbl@provide@lsys{\language}}{}%
3114   \bbl@ifunset{\bbl@wdir@\language}\bbl@provide@dirs{\language}}{}%
3115   \ifx\bbl@mapselect\undefined
3116     \AtBeginDocument{%
3117       \expandafter\bbl@add\csname selectfont \endcsname{\bbl@mapselect}}%
3118       {\selectfont}}%
3119     \def\bbl@mapselect{%
3120       \let\bbl@mapselect\relax
3121       \edef\bbl@prefontid{\fontid\font}}%
3122     \def\bbl@mapdir##1{%
3123       {\def\language{##1}%
3124         \let\bbl@ifrestoring\@firstoftwo % avoid font warning
3125         \bbl@switchfont
3126         \directlua{Babel.fontmap
3127           [\the\csname bbl@wdir@##1\endcsname]
3128           [\bbl@prefontid]=\fontid\font}}}%
3129     \fi
3130     \bbl@exp{\bbl@add\bbl@mapselect{\bbl@mapdir{\language}}}%
3131     \fi
3132   % == Line breaking: intraspace, intrapenalty ==
3133   % For CJK, East Asian, Southeast Asian, if interspace in ini
3134   \ifx\bbl@KVP@intraspace\@nil\else % We can override the ini or set
3135     \bbl@csarg\edef{intsp@#2}{\bbl@KVP@intraspace}%
3136   \fi

```

```

3137 \bbl@provide@intraspace
3138 % == Line breaking: hyphenate.other.locale ==
3139 \bbl@ifunset{bbl@hyotl@languagename}{}%
3140 {\bbl@csarg\bbl@replace{hyotl@languagename}{ }{,}%
3141 \bbl@startcommands*{languagename}{}%
3142 \bbl@csarg\bbl@foreach{hyotl@languagename}{%
3143 \ifcase\bbl@engine
3144 \ifnum##1<257
3145 \SetHyphenMap{BabelLower{##1}{##1}}%
3146 \fi
3147 \else
3148 \SetHyphenMap{BabelLower{##1}{##1}}%
3149 \fi}%
3150 \bbl@endcommands}%
3151 % == Line breaking: hyphenate.other.script ==
3152 \bbl@ifunset{bbl@hyots@languagename}{}%
3153 {\bbl@csarg\bbl@replace{hyots@languagename}{ }{,}%
3154 \bbl@csarg\bbl@foreach{hyots@languagename}{%
3155 \ifcase\bbl@engine
3156 \ifnum##1<257
3157 \global\lccode##1=##1\relax
3158 \fi
3159 \else
3160 \global\lccode##1=##1\relax
3161 \fi}}%
3162 % == Counters: maparabic ==
3163 % Native digits, if provided in ini (TeX level, xe and lua)
3164 \ifcase\bbl@engine\else
3165 \bbl@ifunset{bbl@dgnat@languagename}{}%
3166 {\expandafter\ifx\csname bbl@dgnat@languagename\endcsname\@empty\else
3167 \expandafter\expandafter\expandafter
3168 \bbl@setdigits\csname bbl@dgnat@languagename\endcsname
3169 \ifx\bbl@KVP@maparabic\@nil\else
3170 \ifx\bbl@latinarabic\@undefined
3171 \expandafter\let\expandafter\@arabic
3172 \csname bbl@counter@languagename\endcsname
3173 \else % ie, if layout=counters, which redefines \@arabic
3174 \expandafter\let\expandafter\bbl@latinarabic
3175 \csname bbl@counter@languagename\endcsname
3176 \fi
3177 \fi
3178 \fi}%
3179 \fi
3180 % == Counters: mapdigits ==
3181 % Native digits (lua level).
3182 \ifodd\bbl@engine
3183 \ifx\bbl@KVP@mapdigits\@nil\else
3184 \bbl@ifunset{bbl@dgnat@languagename}{}%
3185 {\RequirePackage{luatexbase}%
3186 \bbl@activate@preotf
3187 \directlua{
3188 Babel = Babel or {} %% -> presets in luababel
3189 Babel.digits_mapped = true
3190 Babel.digits = Babel.digits or {}
3191 Babel.digits[\the\localeid] =
3192 table.pack(string.utfvalue('\bbl@cl{dgnat}'))
3193 if not Babel.numbers then
3194 function Babel.numbers(head)
3195 local LOCALE = luatexbase.registernumber'bbl@attr@locale'

```

```

3196         local GLYPH = node.id'glyph'
3197         local inmath = false
3198         for item in node.traverse(head) do
3199             if not inmath and item.id == GLYPH then
3200                 local temp = node.get_attribute(item, LOCALE)
3201                 if Babel.digits[temp] then
3202                     local chr = item.char
3203                     if chr > 47 and chr < 58 then
3204                         item.char = Babel.digits[temp][chr-47]
3205                     end
3206                 end
3207             elseif item.id == node.id'math' then
3208                 inmath = (item.subtype == 0)
3209             end
3210         end
3211         return head
3212     end
3213 end
3214 }}%
3215 \fi
3216 \fi
3217 % == Counters: alph, Alph ==
3218 % What if extras<lang> contains a \babel@save\@alph? It won't be
3219 % restored correctly when exiting the language, so we ignore
3220 % this change with the \bbl@alph@saved trick.
3221 \ifx\bbl@KVP@alph\@nil\else
3222     \toks@\expandafter\expandafter\expandafter{%
3223         \csname extras\language\endcsname}%
3224     \bbl@exp{%
3225         \def\<extras\language>{%
3226             \let\\bbl@alph@saved\\@alph
3227             \the\toks@
3228             \let\\@alph\\bbl@alph@saved
3229             \\babel@save\\@alph
3230             \let\\@alph<bbl@cntr@\bbl@KVP@alph @\language>}}%
3231 \fi
3232 \ifx\bbl@KVP@Alph\@nil\else
3233     \toks@\expandafter\expandafter\expandafter{%
3234         \csname extras\language\endcsname}%
3235     \bbl@exp{%
3236         \def\<extras\language>{%
3237             \let\\bbl@Alph@saved\\@Alph
3238             \the\toks@
3239             \let\\@Alph\\bbl@Alph@saved
3240             \\babel@save\\@Alph
3241             \let\\@Alph<bbl@cntr@\bbl@KVP@Alph @\language>}}%
3242 \fi
3243 % == require.babel in ini ==
3244 % To load or reload the babel-*.tex, if require.babel in ini
3245 \ifx\bbl@beforestart\relax\else % But not in doc aux or body
3246     \bbl@ifunset{bbl@rqtex@\language}%
3247     {\expandafter\ifx\csname bbl@rqtex@\language\endcsname\@empty\else
3248         \let\BabelBeforeIni\@gobbletwo
3249         \chardef\atcatcode=\catcode`\@
3250         \catcode`\@=11\relax
3251         \bbl@input@texini{\bbl@cs{rqtex@\language}}%
3252         \catcode`\@=\atcatcode
3253         \let\atcatcode\relax
3254     \fi}%

```

```

3255 \fi
3256 % == main ==
3257 \ifx\bbbl@KVP@main\@nil % Restore only if not 'main'
3258 \let\language\bbbl@savelangname
3259 \chardef\localeid\bbbl@savelocaleid\relax
3260 \fi}

```

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

```

3261 \def\bbbl@provide@new#1{%
3262 \namedef{date#1}{}% marks lang exists - required by \StartBabelCommands
3263 \namedef{extras#1}{}%
3264 \namedef{noextras#1}{}%
3265 \bbbl@startcommands*{#1}{captions}%
3266 \ifx\bbbl@KVP@captions\@nil % and also if import, implicit
3267 \def\bbbl@tempb##1{% elt for \bbbl@captionslist
3268 \ifx##1\@empty\else
3269 \bbbl@exp{%
3270 \SetString\##1{%
3271 \bbbl@nocaption{\bbbl@stripslash##1}{#1\bbbl@stripslash##1}}}%
3272 \expandafter\bbbl@tempb
3273 \fi}%
3274 \expandafter\bbbl@tempb\bbbl@captionslist\@empty
3275 \else
3276 \ifx\bbbl@initoload\relax
3277 \bbbl@read@ini{\bbbl@KVP@captions}0% Here letters cat = 11
3278 \else
3279 \bbbl@read@ini{\bbbl@initoload}0% Here all letters cat = 11
3280 \fi
3281 \bbbl@after@ini
3282 \bbbl@savestrings
3283 \fi
3284 \StartBabelCommands*{#1}{date}%
3285 \ifx\bbbl@KVP@import\@nil
3286 \bbbl@exp{%
3287 \SetString\today{\bbbl@nocaption{today}{#1today}}}%
3288 \else
3289 \bbbl@savetoday
3290 \bbbl@savedate
3291 \fi
3292 \bbbl@endcommands
3293 \bbbl@load@basic{#1}%
3294 % == hyphenmins == (only if new)
3295 \bbbl@exp{%
3296 \gdef\<#1hyphenmins>{%
3297 {\bbbl@ifunset{\bbbl@lfthm@#1}{2}{\bbbl@cs{lfthm@#1}}}%
3298 {\bbbl@ifunset{\bbbl@rgthm@#1}{3}{\bbbl@cs{rgthm@#1}}}}}%
3299 % == hyphenrules ==
3300 \bbbl@provide@hyphens{#1}%
3301 % == frenchspacing == (only if new)
3302 \bbbl@ifunset{\bbbl@frspc@#1}{}%
3303 {\edef\bbbl@tempa{\bbbl@cl{frspc}}%
3304 \edef\bbbl@tempa{\expandafter\@car\bbbl@tempa\@nil}%
3305 \if u\bbbl@tempa % do nothing
3306 \else\if n\bbbl@tempa % non french
3307 \expandafter\bbbl@add\csname extras#1\endcsname{%
3308 \let\bbbl@elt\bbbl@fs@elt@i
3309 \bbbl@fs@chars}%
3310 \else\if y\bbbl@tempa % french
3311 \expandafter\bbbl@add\csname extras#1\endcsname{%

```

```

3312         \let\bbl@elt\bbl@fs@elt@ii
3313         \bbl@fs@chars}%
3314     \fi\fi\fi}%
3315 %
3316 \ifx\bbl@KVP@main\@nil\else
3317     \expandafter\main@language\expandafter{#1}%
3318 \fi}
3319 % A couple of macros used above, to avoid hashes #####...
3320 \def\bbl@fs@elt@i#1#2#3{%
3321     \ifnum\sfcode`#1=#2\relax
3322         \babel@savevariable{\sfcode`#1}%
3323         \sfcode`#1=#3\relax
3324     \fi}%
3325 \def\bbl@fs@elt@ii#1#2#3{%
3326     \ifnum\sfcode`#1=#3\relax
3327         \babel@savevariable{\sfcode`#1}%
3328         \sfcode`#1=#2\relax
3329     \fi}%
3330 %
3331 \def\bbl@provide@renew#1{%
3332     \ifx\bbl@KVP@captions\@nil\else
3333         \StartBabelCommands*{#1}{captions}%
3334         \bbl@read@ini{\bbl@KVP@captions}0%   Here all letters cat = 11
3335         \bbl@after@ini
3336         \bbl@savestrings
3337     \EndBabelCommands
3338 \fi
3339 \ifx\bbl@KVP@import\@nil\else
3340     \StartBabelCommands*{#1}{date}%
3341     \bbl@savetoday
3342     \bbl@savedate
3343     \EndBabelCommands
3344 \fi
3345 % == hyphenrules ==
3346 \bbl@provide@hyphens{#1}}
3347 % Load the basic parameters (ids, typography, counters, and a few
3348 % more), while captions and dates are left out. But it may happen some
3349 % data has been loaded before automatically, so we first discard the
3350 % saved values.
3351 \def\bbl@linebreak@export{%
3352     \bbl@exportkey{lnbrk}{typography.linebreaking}{h}%
3353     \bbl@exportkey{hyphr}{typography.hyphenrules}{}%
3354     \bbl@exportkey{lfthm}{typography.lefthyphenmin}{2}%
3355     \bbl@exportkey{rgthm}{typography.righthyphenmin}{3}%
3356     \bbl@exportkey{prehc}{typography.prehyphenchar}{}%
3357     \bbl@exportkey{hyotl}{typography.hyphenate.other.locale}{}%
3358     \bbl@exportkey{hyots}{typography.hyphenate.other.script}{}%
3359     \bbl@exportkey{intsp}{typography.intraspace}{}%
3360     \bbl@exportkey{chrng}{characters.ranges}{}}
3361 \def\bbl@load@basic#1{%
3362     \bbl@ifunset{bbl@inidata@\language}\@nil\else
3363         {\getlocaleproperty\bbl@tempa{\language}{identification/load.level}%
3364         \ifcase\bbl@tempa\else
3365             \bbl@csarg\let{lname@\language}\relax
3366         \fi}%
3367     \bbl@ifunset{bbl@lname@#1}%
3368     {\def\BabelBeforeIni##1##2{%
3369         \begingroup
3370         \let\bbl@ini@captions@aux\@gobbletwo

```



```

3371 \def\bbl@inidate ####1.####2.####3.####4\relax ####5####6}%
3372 \bbl@read@ini{##1}0%
3373 \bbl@linebreak@export
3374 \bbl@exportkey{dgnat}{numbers.digits.native}{}%
3375 \bbl@exportkey{frspc}{typography.frenchspacing}{u}% unset
3376 \ifx\bbl@initoload\relax\endinput\fi
3377 \endgroup}%
3378 \begingroup % boxed, to avoid extra spaces:
3379 \ifx\bbl@initoload\relax
3380 \bbl@input@texini{#1}%
3381 \else
3382 \setbox\z@\hbox{\BabelBeforeIni{\bbl@initoload}{}}%
3383 \fi
3384 \endgroup}%
3385 {}%

```

The hyphenrules option is handled with an auxiliary macro.

```

3386 \def\bbl@provide@hyphens#1{%
3387 \let\bbl@tempa\relax
3388 \ifx\bbl@KVP@hyphenrules\@nil\else
3389 \bbl@replace\bbl@KVP@hyphenrules{ }{,}%
3390 \bbl@foreach\bbl@KVP@hyphenrules{%
3391 \ifx\bbl@tempa\relax % if not yet found
3392 \bbl@ifsamestring{##1}{+}%
3393 {\bbl@exp{\addlanguage<l@##1>}}}%
3394 {}%
3395 \bbl@ifunset{l@##1}%
3396 {}%
3397 {\bbl@exp{\let\bbl@tempa<l@##1>}}}%
3398 \fi}%
3399 \fi
3400 \ifx\bbl@tempa\relax % if no opt or no language in opt found
3401 \ifx\bbl@KVP@import\@nil
3402 \ifx\bbl@initoload\relax\else
3403 \bbl@exp{% and hyphenrules is not empty
3404 \bbl@ifblank{\bbl@cs{hyphr@#1}}}%
3405 {}%
3406 {\let\bbl@tempa<l@bbl@cl{hyphr}>}}}%
3407 \fi
3408 \else % if importing
3409 \bbl@exp{% and hyphenrules is not empty
3410 \bbl@ifblank{\bbl@cs{hyphr@#1}}}%
3411 {}%
3412 {\let\bbl@tempa<l@bbl@cl{hyphr}>}}}%
3413 \fi
3414 \fi
3415 \bbl@ifunset{bbl@tempa}% ie, relax or undefined
3416 {\bbl@ifunset{l@#1}% no hyphenrules found - fallback
3417 {\bbl@exp{\adddialect<l@#1>\language}}}%
3418 {}% so, l@<lang> is ok - nothing to do
3419 {\bbl@exp{\adddialect<l@#1>\bbl@tempa}}}% found in opt list or ini
3420

```

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.

```

3421 \ifx\bbl@readstream\@undefined
3422 \csname newread\endcsname\bbl@readstream
3423 \fi
3424 \def\bbl@input@texini#1{%

```

```

3425 \bbl@bsphack
3426 \bbl@exp{%
3427 \catcode`\\%=14 \catcode`\\|=0
3428 \catcode`\\{=1 \catcode`\\}=2
3429 \lowercase{\InputIfFileExists{babel-#1.tex}{}}}%
3430 \catcode`\\%= \the\catcode`\%\relax
3431 \catcode`\\|= \the\catcode`\|\relax
3432 \catcode`\\{= \the\catcode`\{\relax
3433 \catcode`\\}= \the\catcode`\}\relax}%
3434 \bbl@esphack}
3435 \def\bbl@inipreread#1=#2\@@{%
3436 \bbl@trim@def\bbl@tempa{#1}% Redundant below !!
3437 \bbl@trim\toks@{#2}%
3438 % Move trims here ??
3439 \bbl@ifunset{bbl@KVP@\bbl@section/\bbl@tempa}%
3440 {\bbl@exp{%
3441 \\\g@addto@macro\\bbl@inidata{%
3442 \\\bbl@elt{\bbl@section}{\bbl@tempa}{\the\toks@}}}%
3443 \expandafter\bbl@inireader\bbl@tempa=#2\@@}%
3444 }%
3445 \def\bbl@fetch@ini#1#2{%
3446 \bbl@exp{\def\\bbl@inidata{%
3447 \\\bbl@elt{identification}{tag.ini}{#1}%
3448 \\\bbl@elt{identification}{load.level}{#2}}}%
3449 \openin\bbl@readstream=babel-#1.ini
3450 \ifeof\bbl@readstream
3451 \bbl@error
3452 {There is no ini file for the requested language\\%
3453 (#1). Perhaps you misspelled it or your installation\\%
3454 is not complete.}%
3455 {Fix the name or reinstall babel.}%
3456 \else
3457 \catcode`\[=12 \catcode`\]=12 \catcode`\==12 \catcode`\&=12
3458 \catcode`\;=12 \catcode`\|=12 \catcode`\%=14 \catcode`\-=12
3459 \bbl@info{Importing
3460 \ifcase#2 \or font and identification \or basic \fi
3461 data for \language\name\\%
3462 from babel-#1.ini. Reported}%
3463 \loop
3464 \if \ifeof\bbl@readstream \fi \T\relax % Trick, because inside \loop
3465 \endlinechar\m@ne
3466 \read\bbl@readstream to \bbl@line
3467 \endlinechar`^^M
3468 \ifx\bbl@line\empty\else
3469 \expandafter\bbl@inline\bbl@line\bbl@inline
3470 \fi
3471 \repeat
3472 \fi}
3473 \def\bbl@read@ini#1#2{%
3474 \bbl@csarg\xdef{lini@\language\name}{#1}%
3475 \let\bbl@section\@empty
3476 \let\bbl@savestrings\@empty
3477 \let\bbl@savetoday\@empty
3478 \let\bbl@savestate\@empty
3479 \let\bbl@inireader\bbl@iniskip
3480 \bbl@fetch@ini{#1}{#2}%
3481 \bbl@foreach\bbl@renewlist{%
3482 \bbl@ifunset{bbl@renew@##1}{\bbl@inisec[##1]\@}}}%
3483 \global\let\bbl@renewlist\@empty

```

```

3484 % Ends last section. See \bbl@inisec
3485 \def\bbl@elt##1##2{\bbl@inireader##1=##2\@@}%
3486 \bbl@cs{renew@\bbl@section}%
3487 \global\bbl@csarg\let{renew@\bbl@section}\relax
3488 \bbl@cs{secpost@\bbl@section}%
3489 \bbl@csarg{\global\expandafter\let}{inidata@\language}\bbl@inidata
3490 \bbl@exp{\bbl@add@list\bbl@ini@loaded{\language}}%
3491 \bbl@to\global\bbl@ini@loaded}
3492 \def\bbl@inline#1\bbl@inline{%
3493 \@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.

```

3494 \def\bbl@iniskip#1\@@{%          if starts with ;
3495 \def\bbl@inisec[#1]#2\@@{%      if starts with opening bracket
3496 \def\bbl@elt##1##2{%
3497   \expandafter\toks@\expandafter{%
3498     \expandafter{\bbl@section}{##1}{##2}}%
3499   \bbl@exp{%
3500     \g@addto@macro\bbl@inidata{\bbl@elt\the\toks@}}%
3501   \bbl@inireader##1=##2\@@}%
3502 \bbl@cs{renew@\bbl@section}%
3503 \global\bbl@csarg\let{renew@\bbl@section}\relax
3504 \bbl@cs{secpost@\bbl@section}%
3505 % The previous code belongs to the previous section.
3506 % -----
3507 % Now start the current one.
3508 \in@{=date.}{#1}%
3509 \ifin@
3510   \lowercase{\def\bbl@tempa{#1=}}%
3511   \bbl@replace\bbl@tempa{=date.gregorian.}{}%
3512   \bbl@replace\bbl@tempa{=date.}{}%
3513   \in@{.licr=}{#1=}%
3514   \ifin@
3515     \ifcase\bbl@engine
3516       \bbl@replace\bbl@tempa{.licr=}{}%
3517     \else
3518       \let\bbl@tempa\relax
3519     \fi
3520   \fi
3521   \ifx\bbl@tempa\relax\else
3522     \bbl@replace\bbl@tempa{=}{}%
3523     \bbl@exp{%
3524       \def<\bbl@inikv@#1>####1=####2\@@{%
3525         \bbl@inidate####1...\relax{####2}{\bbl@tempa}}%
3526       \fi
3527     \fi
3528   \def\bbl@section{#1}%
3529   \def\bbl@elt##1##2{%
3530     \@namedef{\bbl@KVP@#1/#1}{}}%
3531   \bbl@cs{renew@#1}%
3532   \bbl@cs{secpre@#1}% pre-section `hook'
3533   \bbl@ifunset{\bbl@inikv@#1}%
3534     {\let\bbl@inireader\bbl@iniskip}%
3535     {\bbl@exp{\let\bbl@inireader<\bbl@inikv@#1>}}
3536 \let\bbl@renewlist\@empty
3537 \def\bbl@renewinikey#1/#2\@@#3{%

```

```

3538 \bbl@ifunset{\bbl@renew@#1}%
3539 {\bbl@add@list\bbl@renewlist{#1}}%
3540 {}%
3541 \bbl@csarg\bbl@add{renew@#1}{\bbl@elt{#2}{#3}}

```

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

```

3542 \def\bbl@inikv#1=#2\@{%      key=value
3543 \bbl@trim@def\bbl@tempa{#1}%
3544 \bbl@trim\toks@{#2}%
3545 \bbl@csarg\edef{\bbl@section.\bbl@tempa}{\the\toks@}}

```

The previous assignments are local, so we need to export them. If the value is empty, we can provide a default value.

```

3546 \def\bbl@exportkey#1#2#3{%
3547 \bbl@ifunset{\bbl@kv@#2}%
3548 {\bbl@csarg\gdef{#1@\language}\language{#3}}%
3549 {\expandafter\ifx\csname\bbl@kv@#2\endcsname\@empty
3550 \bbl@csarg\gdef{#1@\language}\language{#3}}%
3551 \else
3552 \bbl@exp{\global\let\<bbl@#1@\language>\<bbl@kv@#2>}%
3553 \fi}}

```

Key-value pairs are treated differently depending on the section in the ini file. The following macros are the readers for identification and typography. Note \bbl@secpost@identification is called always (via \bbl@inisec), while \bbl@after@ini must be called explicitly after \bbl@read@ini if necessary.

```

3554 \def\bbl@iniwarning#1{%
3555 \bbl@ifunset{\bbl@kv@identification.warning#1}{}%
3556 {\bbl@warning%
3557 From babel-\bbl@cs{lini@\language}.ini:\%
3558 \bbl@cs{@kv@identification.warning#1}\%
3559 Reported }}%
3560 %
3561 \let\bbl@inikv@identification\bbl@inikv
3562 \def\bbl@secpost@identification{%
3563 \bbl@iniwarning}%
3564 \ifcase\bbl@engine
3565 \bbl@iniwarning{.pdflatex}%
3566 \or
3567 \bbl@iniwarning{.lualatex}%
3568 \or
3569 \bbl@iniwarning{.xelatex}%
3570 \fi%
3571 \bbl@exportkey{elname}{identification.name.english}{}%
3572 \bbl@exp{\bbl@exportkey{lname}{identification.name.opentype}%
3573 {\csname\bbl@elname@\language\endcsname}}%
3574 \bbl@exportkey{tbcp}{identification.tag.bcp47}{}%
3575 \bbl@exportkey{lbcpl}{identification.language.tag.bcp47}{}%
3576 \bbl@exportkey{lotf}{identification.tag.opentype}{dflt}%
3577 \bbl@exportkey{esname}{identification.script.name}{}%
3578 \bbl@exp{\bbl@exportkey{sname}{identification.script.name.opentype}%
3579 {\csname\bbl@esname@\language\endcsname}}%
3580 \bbl@exportkey{sbcpl}{identification.script.tag.bcp47}{}%
3581 \bbl@exportkey{sotf}{identification.script.tag.opentype}{DFLT}%
3582 \ifbbl@bcptoname
3583 \bbl@csarg\xdef{bcp@map@\bbl@cl{tbcp}}{\language}%
3584 \fi}

```

By default, the following sections are just read. Actions are taken later.

```

3585 \let\bbl@inikv@typography\bbl@inikv
3586 \let\bbl@inikv@characters\bbl@inikv
3587 \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 \localenumeral, and another one preserving the trailing .1 for the ‘units’.

```

3588 \def\bbl@inikv@counters#1=#2\@@{%
3589   \bbl@ifsamestring{#1}{digits}%
3590   {\bbl@error{The counter name 'digits' is reserved for mapping\\%
3591     decimal digits}%
3592     {Use another name.}}%
3593   }%
3594 \def\bbl@tempc{#1}%
3595 \bbl@trim@def{\bbl@tempb*}{#2}%
3596 \in@{.1$}{#1$}%
3597 \ifin@
3598   \bbl@replace\bbl@tempc{.1}{}%
3599   \bbl@csarg\protected@xdef{cntr@\bbl@tempc @\language}\bbl@tempc
3600   \noexpand\bbl@alphanumeric{\bbl@tempc}}%
3601 \fi
3602 \in@{.F.}{#1}%
3603 \ifin@else\in@{.S.}{#1}\fi
3604 \ifin@
3605   \bbl@csarg\protected@xdef{cntr@#1@\language}\bbl@tempb*}%
3606 \else
3607   \toks@{}% Required by \bbl@buildifcase, which returns \bbl@tempa
3608   \expandafter\bbl@buildifcase\bbl@tempb* \ \ % Space after \
3609   \bbl@csarg{\global\expandafter\let}{cntr@#1@\language}\bbl@tempa
3610 \fi}
3611 \def\bbl@after@ini{%
3612   \bbl@linebreak@export
3613   \bbl@exportkey{dgnat}{numbers.digits.native}{}%
3614   \bbl@exportkey{rqtex}{identification.require.babel}{}%
3615   \bbl@exportkey{frspc}{typography.frenchspacing}{u}% unset
3616   \bbl@toglobal\bbl@savetoday
3617   \bbl@toglobal\bbl@savestate}

```

Now captions and captions.licr, depending on the engine. And below also for dates. They rely on a few auxiliary macros. It is expected the ini file provides the complete set in Unicode and LICR, in that order.

```

3618 \ifcase\bbl@engine
3619   \bbl@csarg\def{inikv@captions.licr}#1=#2\@@{%
3620     \bbl@ini@captions@aux{#1}{#2}}
3621 \else
3622   \def\bbl@inikv@captions#1=#2\@@{%
3623     \bbl@ini@captions@aux{#1}{#2}}
3624 \fi

```

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

```

3625 \def\bbl@ini@captions@aux#1#2{%
3626   \bbl@trim@def\bbl@tempa{#1}%
3627   \bbl@xin@{.template}{\bbl@tempa}%
3628   \ifin@
3629     \bbl@replace\bbl@tempa{.template}{}%
3630   \def\bbl@toreplace{#2}%
3631   \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace}}%
3632   \bbl@replace\bbl@toreplace{[ ]}{\csname}%
3633   \bbl@replace\bbl@toreplace{[ ]}{\csname the}%

```

```

3634 \bbl@replace\bbl@toreplace{}}{name\endcsname{}}}%
3635 \bbl@replace\bbl@toreplace{}}{\endcsname{}}}%
3636 \bbl@xin@{,\bbl@tempa,}{,chapter,}%
3637 \ifin@
3638 \bbl@patchchapter
3639 \global\bbl@csarg\let{chapfmt@\language}\bbl@toreplace
3640 \fi
3641 \bbl@xin@{,\bbl@tempa,}{,appendix,}%
3642 \ifin@
3643 \bbl@patchchapter
3644 \global\bbl@csarg\let{appxfmt@\language}\bbl@toreplace
3645 \fi
3646 \bbl@xin@{,\bbl@tempa,}{,part,}%
3647 \ifin@
3648 \bbl@patchpart
3649 \global\bbl@csarg\let{partfmt@\language}\bbl@toreplace
3650 \fi
3651 \bbl@xin@{,\bbl@tempa,}{,figure,table,}%
3652 \ifin@
3653 \toks@\expandafter{\bbl@toreplace}%
3654 \bbl@exp{\gdef\<fnum@\bbl@tempa>{\the\toks@}}%
3655 \fi
3656 \else
3657 \bbl@ifblank{#2}%
3658 {\bbl@exp%
3659 \toks@{\bbl@nocaption{\bbl@tempa}{\language\bbl@tempa name}}}%
3660 {\bbl@trim\toks@{#2}}%
3661 \bbl@exp%
3662 \bbl@add\bbl@savestrings%
3663 \SetString\<\bbl@tempa name>{\the\toks@}}%
3664 \toks@\expandafter{\bbl@captionslist}%
3665 \bbl@exp{\in{\<\bbl@tempa name>}{\the\toks@}}%
3666 \ifin@else
3667 \bbl@exp%
3668 \bbl@add\<\bbl@extracaps@\language>{\<\bbl@tempa name>}%
3669 \bbl@to\global\<\bbl@extracaps@\language>}%
3670 \fi
3671 \fi}

```

Labels. Captions must contain just strings, no format at all, so there is new group in ini files.

```

3672 \def\bbl@list@the{%
3673 part,chapter,section,subsection,subsubsection,paragraph,%
3674 subparagraph,enumi,enumii,enumiii,enumiv,equation,figure,%
3675 table,page,footnote,mpfootnote,mpfn}
3676 \def\bbl@map@cnt#1{% #1:roman,etc, // #2:enumi,etc
3677 \bbl@ifunset{bbl@map@#1@\language}%
3678 {\nameuse{#1}}%
3679 {\nameuse{bbl@map@#1@\language}}}
3680 \def\bbl@inikv@labels#1=#2\@@{%
3681 \in@{.map}{#1}}%
3682 \ifin@
3683 \ifx\bbl@KVP@labels\@nil\else
3684 \bbl@xin@{ map }{ \bbl@KVP@labels\space}%
3685 \ifin@
3686 \def\bbl@tempc{#1}%
3687 \bbl@replace\bbl@tempc{.map}{}%
3688 \in@{,#2,}{,arabic,roman,Roman,alph,Alph,fnsymbol,}%
3689 \bbl@exp%

```

```

3690 \gdef\<bbl@map@bbl@tempc @\language>%
3691 {\ifin@<#2>\else\\localecounter{#2}\fi}}%
3692 \bbl@foreach\bbl@list@the{%
3693 \bbl@ifunset{the##1}{}%
3694 {\bbl@exp{\let\\bbl@tempd\<the##1>}%
3695 \bbl@exp{%
3696 \\bbl@sreplace\<the##1>%
3697 {\<bbl@tempc>{##1}}{\\bbl@map@cnt{\bbl@tempc}{##1}}%
3698 \\bbl@sreplace\<the##1>%
3699 {\<\@empty @\bbl@tempc>\<c@##1>}{\\bbl@map@cnt{\bbl@tempc}{##1}}}%
3700 \expandafter\ifx\csname the##1\endcsname\bbl@tempd\else
3701 \toks@ \expandafter\expandafter\expandafter{%
3702 \csname the##1\endcsname}%
3703 \expandafter\xdef\csname the##1\endcsname{{\the\toks@}}%
3704 \fi}}%
3705 \fi
3706 \fi
3707 %
3708 \else
3709 %
3710 % The following code is still under study. You can test it and make
3711 % suggestions. Eg, enumerate.2 = ([enumi]).([enumii]). It's
3712 % language dependent.
3713 \in@{enumerate.}{#1}%
3714 \ifin@
3715 \def\bbl@tempa{#1}%
3716 \bbl@replace\bbl@tempa{enumerate.}{}%
3717 \def\bbl@toreplace{#2}%
3718 \bbl@replace\bbl@toreplace{[ ]}{\nobreakspace{}}%
3719 \bbl@replace\bbl@toreplace{[ ]}{\csname the}%
3720 \bbl@replace\bbl@toreplace{ ]}{\endcsname{}}}%
3721 \toks@ \expandafter{\bbl@toreplace}%
3722 \bbl@exp{%
3723 \\bbl@add\<extras\language>{%
3724 \\babel@save\<labelenum\romannumeral\bbl@tempa>%
3725 \def\<labelenum\romannumeral\bbl@tempa>{\the\toks@}}%
3726 \\bbl@tglobal\<extras\language>}%
3727 \fi
3728 \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.

```

3729 \def\bbl@chapttype{chap}
3730 \ifx\@makechapterhead\@undefined
3731 \let\bbl@patchchapter\relax
3732 \else\ifx\thechapter\@undefined
3733 \let\bbl@patchchapter\relax
3734 \else\ifx\ps@headings\@undefined
3735 \let\bbl@patchchapter\relax
3736 \else
3737 \def\bbl@patchchapter{%
3738 \global\let\bbl@patchchapter\relax
3739 \bbl@add\appendix{\def\bbl@chapttype{appx}}% Not harmful, I hope
3740 \bbl@tglobal\appendix
3741 \bbl@sreplace\ps@headings
3742 {\@chapapp\ thechapter}%
3743 {\bbl@chapterformat}%

```

```

3744 \bbl@toglobal\ps@headings
3745 \bbl@sreplace\chaptermark
3746 {\@chapapp\ thechapter}%
3747 {\bbl@chapterformat}%
3748 \bbl@toglobal\chaptermark
3749 \bbl@sreplace\@makechapterhead
3750 {\@chapapp\space\thechapter}%
3751 {\bbl@chapterformat}%
3752 \bbl@toglobal\@makechapterhead
3753 \gdef\bbl@chapterformat{%
3754 \bbl@ifunset{\bbl@bbl@chapttype fmt@\language}%
3755 {\@chapapp\space\thechapter}
3756 {\@nameuse{\bbl@bbl@chapttype fmt@\language}}}}
3757 \fi\fi\fi
3758 \ifx\@part\@undefined
3759 \let\bbl@patchpart\relax
3760 \else
3761 \def\bbl@patchpart{%
3762 \global\let\bbl@patchpart\relax
3763 \bbl@sreplace\@part
3764 {\partname\nobreakspace\thepart}%
3765 {\bbl@partformat}%
3766 \bbl@toglobal\@part
3767 \gdef\bbl@partformat{%
3768 \bbl@ifunset{\bbl@partfmt@\language}%
3769 {\partname\nobreakspace\thepart}
3770 {\@nameuse{\bbl@partfmt@\language}}}}
3771 \fi

Date. TODO. Document

3772 % Arguments are _not_ protected.
3773 \let\bbl@calendar\@empty
3774 \DeclareRobustCommand\localedate[1][\bbl@localedate{#1}]
3775 \def\bbl@localedate#1#2#3#4{%
3776 \begin{group}
3777 \ifx\@empty#1\@empty\else
3778 \let\bbl@ld@calendar\@empty
3779 \let\bbl@ld@variant\@empty
3780 \edef\bbl@tempa{\zap@space#1 \@empty}%
3781 \def\bbl@tempb##1=##2\@{\@namedef{\bbl@ld@##1}{##2}}%
3782 \bbl@foreach\bbl@tempa{\bbl@tempb##1\@}%
3783 \edef\bbl@calendar{%
3784 \bbl@ld@calendar
3785 \ifx\bbl@ld@variant\@empty\else
3786 .\bbl@ld@variant
3787 \fi}%
3788 \bbl@replace\bbl@calendar{gregorian}{}}%
3789 \fi
3790 \bbl@cased
3791 {\@nameuse{\bbl@date@\language @\bbl@calendar}{#2}{#3}{#4}}%
3792 \end{group}
3793 % eg: 1=months, 2=wide, 3=1, 4=dummy, 5=value, 6=calendar
3794 \def\bbl@inidate#1.#2.#3.#4\relax#5#6{% TODO - ignore with 'captions'
3795 \bbl@trim@def\bbl@tempa{#1.#2}%
3796 \bbl@ifsamestring{\bbl@tempa}{months.wide}% to savedate
3797 {\bbl@trim@def\bbl@tempa{#3}%
3798 \bbl@trim\toks@{#5}%
3799 \@temptokena\expandafter{\bbl@savedate}%
3800 \bbl@exp{% Reverse order - in ini last wins

```



```

3801 \def\\bbl@savestate{%
3802   \\SetString\<month\romannumeral\bbl@tempa#6name>\the\toks@}%
3803   \the\temptokena}}}%
3804 {\bbl@ifsamestring{\bbl@tempa}{date.long}%      defined now
3805   {\lowercase{\def\bbl@tempb{#6}}}%
3806   \bbl@trim@def\bbl@toreplace{#5}%
3807   \bbl@TG@date
3808   \bbl@ifunset{\bbl@date@\language @}%
3809   {\global\bbl@csarg\let{date@\language @}\bbl@toreplace
3810   % TODO. Move to a better place.
3811   \bbl@exp{%
3812     \gdef\<\language date>\protect\<\language date >}%
3813     \gdef\<\language date >####1####2####3{%
3814       \\bbl@usedategroupttrue
3815       \<bbl@ensure@\language>{%
3816         \\localedate{####1}{####2}{####3}}}%
3817       \\bbl@add\\bbl@savetoday{%
3818         \\SetString\\today{%
3819           \<\language date>%
3820           {\the\year}{\the\month}{\the\day}}}%
3821       }%
3822   \ifx\bbl@tempb\empty\else
3823     \global\bbl@csarg\let{date@\language @\bbl@tempb}\bbl@toreplace
3824   \fi}%
3825   {}}

```

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.

```

3826 \let\bbl@calendar\empty
3827 \newcommand\BabelDateSpace{\nobreakspace}
3828 \newcommand\BabelDateDot{.\@} % TODO. \let instead of repeating
3829 \newcommand\BabelDated[1]{\number#1}
3830 \newcommand\BabelDatedd[1]{\ifnum#1<10 0\fi\number#1}
3831 \newcommand\BabelDateM[1]{\number#1}
3832 \newcommand\BabelDateMM[1]{\ifnum#1<10 0\fi\number#1}
3833 \newcommand\BabelDateMMM[1]{%
3834   \csname month\romannumeral#1\bbl@calendar name\endcsname}}%
3835 \newcommand\BabelDatey[1]{\number#1}%
3836 \newcommand\BabelDateyy[1]{%
3837   \ifnum#1<10 0\number#1 %
3838   \else\ifnum#1<100 \number#1 %
3839   \else\ifnum#1<1000 \expandafter\@gobble\number#1 %
3840   \else\ifnum#1<10000 \expandafter\@gobbletwo\number#1 %
3841   \else
3842     \bbl@error
3843     {Currently two-digit years are restricted to the\
3844     range 0-9999.}%
3845     {There is little you can do. Sorry.}%
3846   \fi\fi\fi\fi}}
3847 \newcommand\BabelDateyyyy[1]{\number#1} % FIXME - add leading 0
3848 \def\bbl@replace@finish@iii#1{%
3849   \bbl@exp{\def\#1####1####2####3{\the\toks@}}
3850 \def\bbl@TG@date{%
3851   \bbl@replace\bbl@toreplace{[ ]}{\BabelDateSpace}}%
3852   \bbl@replace\bbl@toreplace{[. ]}{\BabelDateDot}}%
3853   \bbl@replace\bbl@toreplace{[d]}{\BabelDated{####3}}%
3854   \bbl@replace\bbl@toreplace{[dd]}{\BabelDatedd{####3}}%
3855   \bbl@replace\bbl@toreplace{[M]}{\BabelDateM{####2}}%

```

```

3856 \bbl@replace\bbl@toreplace{[MM]}\BabelDateMM{####2}%
3857 \bbl@replace\bbl@toreplace{[MMMM]}\BabelDateMMMM{####2}%
3858 \bbl@replace\bbl@toreplace{[y]}\BabelDatey{####1}%
3859 \bbl@replace\bbl@toreplace{[yy]}\BabelDateyy{####1}%
3860 \bbl@replace\bbl@toreplace{[yyyy]}\BabelDateyyyy{####1}%
3861 \bbl@replace\bbl@toreplace{[y]}\bbl@datecctr[####1]|}%
3862 \bbl@replace\bbl@toreplace{[m]}\bbl@datecctr[####2]|}%
3863 \bbl@replace\bbl@toreplace{[d]}\bbl@datecctr[####3]|}%
3864 % Note after \bbl@replace \toks@ contains the resulting string.
3865 % TODO - Using this implicit behavior doesn't seem a good idea.
3866 \bbl@replace@finish@iii\bbl@toreplace}
3867 \def\bbl@datecctr\expandafter\bbl@xdatecctr\expandafter}
3868 \def\bbl@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.

```

3869 \def\bbl@provide@lsys#1{%
3870   \bbl@ifunset{bbl@lname@#1}%
3871     {\bbl@ini@basic{#1}}%
3872     {}%
3873   \bbl@csarg\let{lsys@#1}\empty
3874   \bbl@ifunset{bbl@sname@#1}\bbl@csarg\gdef{sname@#1}{Default}}}%
3875   \bbl@ifunset{bbl@sotf@#1}\bbl@csarg\gdef{sotf@#1}{DFLT}}}%
3876   \bbl@csarg\bbl@add@list{lsys@#1}{Script=\bbl@cs{sname@#1}}}%
3877   \bbl@ifunset{bbl@lname@#1}{%
3878     {\bbl@csarg\bbl@add@list{lsys@#1}{Language=\bbl@cs{lname@#1}}}%
3879   \ifcase\bbl@engine\or\or
3880     \bbl@ifunset{bbl@prehc@#1}{%
3881       {\bbl@exp{\bbl@ifblank{\bbl@cs{prehc@#1}}}%
3882       {}%
3883       {\ifx\bbl@xenohyph\undefined
3884         \let\bbl@xenohyph\bbl@xenohyph@d
3885         \ifx\AtBeginDocument\@notprerr
3886           \expandafter\@secondoftwo % to execute right now
3887           \fi
3888         \AtBeginDocument{%
3889           \expandafter\bbl@add
3890           \csname selectfont \endcsname{\bbl@xenohyph}%
3891           \expandafter\selectlanguage\expandafter{\languagename}%
3892           \expandafter\bbl@tglobal\csname selectfont \endcsname}%
3893         \fi}}%
3894     \fi
3895   \bbl@csarg\bbl@tglobal{lsys@#1}}
3896 \def\bbl@xenohyph@d{%
3897   \bbl@ifset{bbl@prehc@languagename}%
3898     {\ifnum\hyphenchar\font=\defaultshyphenchar
3899       \iffontchar\font\bbl@cl{prehc}\relax
3900       \hyphenchar\font\bbl@cl{prehc}\relax
3901       \else\iffontchar\font"200B
3902         \hyphenchar\font"200B
3903       \else
3904         \bbl@warning
3905         {Neither 0 nor ZERO WIDTH SPACE are available\\%
3906         in the current font, and therefore the hyphen\\%
3907         will be printed. Try changing the fontspec's\\%
3908         'HyphenChar' to another value, but be aware\\%
3909         this setting is not safe (see the manual)}%
3910         \hyphenchar\font\defaultshyphenchar
3911       \fi\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).

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 T_EX. Non-digits characters are kept. The first macro is the generic “localized” command.

Alphabetic counters must be converted from a space separated list to an \ifcase structure.

154

```

3958 \<ifcase>###1\space\the\toks@\<else>\\\@ctrerr\<fi>}}%
3959 \else
3960 \toks@\expandafter{\the\toks@\or #1}%
3961 \expandafter\bb1@buildifcase
3962 \fi}

```

The code for additive counters is somewhat tricky and it's based on the fact the arguments just before @@ collect digits which have been left 'unused' in previous arguments, the first of them being the number of digits in the number to be converted. This explains the reverse set 76543210. Digits above 10000 are not handled yet. When the key contains the subkey .F., the number after is treated as a special case, for a fixed form (see babel-he.ini, for example).

```

3963 \newcommand\localenumeral[2]{\bb1@cs{cntr@#1@\language}\{#2}}
3964 \def\bb1@localecntr#1#2{\localenumeral{#2}{#1}}
3965 \newcommand\localecounter[2]{%
3966 \expandafter\bb1@localecntr
3967 \expandafter{\number\csname c@#2\endcsname}\{#1}}
3968 \def\bb1@alphnumeral#1#2{%
3969 \expandafter\bb1@alphnumeral@i\number#2 76543210\@{#1}}
3970 \def\bb1@alphnumeral@i#1#2#3#4#5#6#7#8\@#9{%
3971 \ifcase\car#8\@nil\or % Currently <10000, but prepared for bigger
3972 \bb1@alphnumeral@ii{#9}000000#1\or
3973 \bb1@alphnumeral@ii{#9}00000#1#2\or
3974 \bb1@alphnumeral@ii{#9}0000#1#2#3\or
3975 \bb1@alphnumeral@ii{#9}000#1#2#3#4\else
3976 \bb1@alphnum@invalid{>9999}%
3977 \fi}
3978 \def\bb1@alphnumeral@ii#1#2#3#4#5#6#7#8{%
3979 \bb1@ifunset{bb1@cntr@#1.F.\number#5#6#7#8@\language}%
3980 {\bb1@cs{cntr@#1.4@\language}\{#5}
3981 \bb1@cs{cntr@#1.3@\language}\{#6}
3982 \bb1@cs{cntr@#1.2@\language}\{#7}
3983 \bb1@cs{cntr@#1.1@\language}\{#8}
3984 \ifnum#6#7#8>\z@ % TODO. An ad hoc rule for Greek. Ugly.
3985 \bb1@ifunset{bb1@cntr@#1.S.321@\language}\{}}%
3986 {\bb1@cs{cntr@#1.S.321@\language}\{}}%
3987 \fi}%
3988 {\bb1@cs{cntr@#1.F.\number#5#6#7#8@\language}}%
3989 \def\bb1@alphnum@invalid#1{%
3990 \bb1@error{Alphabetic numeral too large (#1)}%
3991 {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.

```

3992 \newcommand\localeinfo[1]{%
3993 \bb1@ifunset{bb1@csname bb1@info@#1\endcsname @\language}%
3994 {\bb1@error{I've found no info for the current locale.\%
3995 The corresponding ini file has not been loaded\%
3996 Perhaps it doesn't exist}%
3997 {See the manual for details.}}%
3998 {\bb1@cs{\csname bb1@info@#1\endcsname @\language}}%
3999 % \@namedef{bb1@info@name.locale}\{lname}
4000 \@namedef{bb1@info@tag.ini}\{lini}
4001 \@namedef{bb1@info@name.english}\{elname}
4002 \@namedef{bb1@info@name.opentype}\{lname}
4003 \@namedef{bb1@info@tag.bcp47}\{tbc}
4004 \@namedef{bb1@info@language.tag.bcp47}\{lbc}
4005 \@namedef{bb1@info@tag.opentype}\{lotf}

```

```

4006 \@namedef{bbl@info@script.name}{esname}
4007 \@namedef{bbl@info@script.name.opentype}{sname}
4008 \@namedef{bbl@info@script.tag.bcp47}{sbcp}
4009 \@namedef{bbl@info@script.tag.opentype}{sotf}
4010 \let\bbl@ensureinfo\@gobble
4011 \newcommand\BabelEnsureInfo{%
4012   \ifx\InputIfFileExists\undefined\else
4013     \def\bbl@ensureinfo##1{%
4014       \bbl@ifunset{bbl@lname@##1}{\bbl@ini@basic{##1}}{}}%
4015   \fi
4016   \bbl@foreach\bbl@loaded{%
4017     \def\language{##1}%
4018     \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`.

```

4019 \newcommand\getlocaleproperty{%
4020   \@ifstar\bbl@getproperty@s\bbl@getproperty@x}
4021 \def\bbl@getproperty@s#1#2#3{%
4022   \let#1\relax
4023   \def\bbl@elt##1##2##3{%
4024     \bbl@ifsamestring{##1/##2}{#3}%
4025     {\providecommand#1{##3}%
4026     \def\bbl@elt####1####2####3{}}}%
4027   {}}%
4028   \bbl@cs{inidata@#2}}%
4029 \def\bbl@getproperty@x#1#2#3{%
4030   \bbl@getproperty@s{#1}{#2}{#3}%
4031   \ifx#1\relax
4032     \bbl@error
4033     {Unknown key for locale '#2':\%
4034     #3\%
4035     \string#1 will be set to \relax}%
4036     {Perhaps you misspelled it.}%
4037   \fi}
4038 \let\bbl@ini@loaded\@empty
4039 \newcommand\LocaleForEach{\bbl@foreach\bbl@ini@loaded}

```

10 Adjusting the Babel behavior

A generic high level interface is provided to adjust some global and general settings.

```

4040 \newcommand\babeladjust[1]{% TODO. Error handling.
4041   \bbl@forkv{#1}{%
4042     \bbl@ifunset{bbl@ADJ@##1@##2}%
4043     {\bbl@cs{ADJ@##1}{##2}}%
4044     {\bbl@cs{ADJ@##1@##2}}}
4045 %
4046 \def\bbl@adjust@lua#1#2{%
4047   \ifvmode
4048     \ifnum\currentgrouplevel=\z@
4049       \directlua{ Babel.#2 }%
4050       \expandafter\expandafter\expandafter\@gobble
4051     \fi
4052   \fi
4053   {\bbl@error % The error is gobbled if everything went ok.
4054     {Currently, #1 related features can be adjusted only\%

```

```

4055         in the main vertical list.}%
4056         {Maybe things change in the future, but this is what it is.}}
4057 \@namedef{bbl@ADJ@bidi.mirroring@on}{%
4058   \bbl@adjust@lua{bidi}{mirroring_enabled=true}}
4059 \@namedef{bbl@ADJ@bidi.mirroring@off}{%
4060   \bbl@adjust@lua{bidi}{mirroring_enabled=false}}
4061 \@namedef{bbl@ADJ@bidi.text@on}{%
4062   \bbl@adjust@lua{bidi}{bidi_enabled=true}}
4063 \@namedef{bbl@ADJ@bidi.text@off}{%
4064   \bbl@adjust@lua{bidi}{bidi_enabled=false}}
4065 \@namedef{bbl@ADJ@bidi.mapdigits@on}{%
4066   \bbl@adjust@lua{bidi}{digits_mapped=true}}
4067 \@namedef{bbl@ADJ@bidi.mapdigits@off}{%
4068   \bbl@adjust@lua{bidi}{digits_mapped=false}}
4069 %
4070 \@namedef{bbl@ADJ@linebreak.sea@on}{%
4071   \bbl@adjust@lua{linebreak}{sea_enabled=true}}
4072 \@namedef{bbl@ADJ@linebreak.sea@off}{%
4073   \bbl@adjust@lua{linebreak}{sea_enabled=false}}
4074 \@namedef{bbl@ADJ@linebreak.cjk@on}{%
4075   \bbl@adjust@lua{linebreak}{cjk_enabled=true}}
4076 \@namedef{bbl@ADJ@linebreak.cjk@off}{%
4077   \bbl@adjust@lua{linebreak}{cjk_enabled=false}}
4078 %
4079 \def\bbl@adjust@layout#1{%
4080   \ifvmode
4081     #1%
4082     \expandafter\@gobble
4083   \fi
4084   {\bbl@error   % The error is gobbled if everything went ok.
4085     {Currently, layout related features can be adjusted only\\%
4086       in vertical mode.}%
4087     {Maybe things change in the future, but this is what it is.}}
4088 \@namedef{bbl@ADJ@layout.tabular@on}{%
4089   \bbl@adjust@layout{\let\@tabular\bbl@NL@tabular}}
4090 \@namedef{bbl@ADJ@layout.tabular@off}{%
4091   \bbl@adjust@layout{\let\@tabular\bbl@OL@tabular}}
4092 \@namedef{bbl@ADJ@layout.lists@on}{%
4093   \bbl@adjust@layout{\let\list\bbl@NL@list}}
4094 \@namedef{bbl@ADJ@layout.lists@off}{%
4095   \bbl@adjust@layout{\let\list\bbl@OL@list}}
4096 \@namedef{bbl@ADJ@hyphenation.extra@on}{%
4097   \bbl@activateposthyphen}
4098 %
4099 \@namedef{bbl@ADJ@autoload.bcp47@on}{%
4100   \bbl@bcpallowedtrue}
4101 \@namedef{bbl@ADJ@autoload.bcp47@off}{%
4102   \bbl@bcpallowedfalse}
4103 \@namedef{bbl@ADJ@autoload.bcp47.prefix}#1{%
4104   \def\bbl@bcp@prefix{#1}}
4105 \def\bbl@bcp@prefix{bcp47-}
4106 \@namedef{bbl@ADJ@autoload.options}#1{%
4107   \def\bbl@autoload@options{#1}}
4108 \let\bbl@autoload@bcptoptions\@empty
4109 \@namedef{bbl@ADJ@autoload.bcp47.options}#1{%
4110   \def\bbl@autoload@bcptoptions{#1}}
4111 \newif\ifbbl@bcptoname
4112 \@namedef{bbl@ADJ@bcp47.toname@on}{%
4113   \bbl@bcptonametrue

```

```

4114 \BabelEnsureInfo}
4115 \@namedef{bbl@ADJ@bcp47.toname@off}{%
4116 \bbl@bcptonamefalse}
4117% TODO: use babel name, override
4118%
4119% As the final task, load the code for lua.
4120%
4121 \ifx\directlua\@undefined\else
4122 \ifx\bbl@luapatterns\@undefined
4123 \input luababel.def
4124 \fi
4125 \fi
4126 \</core>

A proxy file for switch.def

4127 \<*kernel>
4128 \let\bbl@onlyswitch\@empty
4129 \input babel.def
4130 \let\bbl@onlyswitch\@undefined
4131 \</kernel>
4132 \<*patterns>

```

11 Loading hyphenation patterns

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

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

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

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

```

4133 \<<Make sure ProvidesFile is defined>>
4134 \ProvidesFile{hyphen.cfg}[\<<date>>] [\<<version>>] Babel hyphens]
4135 \xdef\bbl@format{\jobname}
4136 \def\bbl@version{\<<version>>}
4137 \def\bbl@date{\<<date>>}
4138 \ifx\AtBeginDocument\@undefined
4139 \def\@empty{}
4140 \let\orig@dump\dump
4141 \def\dump{%
4142 \ifx\@ztryfc\@undefined
4143 \else
4144 \toks0=\expandafter{\@preamblecmds}%
4145 \edef\@preamblecmds{\noexpand\@begindocumenthook\the\toks0}%
4146 \def\@begindocumenthook{}%
4147 \fi
4148 \let\dump\orig@dump\let\orig@dump\@undefined\dump}
4149 \fi
4150 \<<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.

```
4151 \def\process@line#1#2 #3 #4 {%
4152   \ifx=#1%
4153     \process@synonym{#2}%
4154   \else
4155     \process@language{#1#2}{#3}{#4}%
4156   \fi
4157   \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.

```
4158 \toks@{}
4159 \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.

```
4160 \def\process@synonym#1{%
4161   \ifnum\last@language=\m@ne
4162     \toks@\expandafter{\the\toks@\relax\process@synonym{#1}}%
4163   \else
4164     \expandafter\chardef\csname l@#1\endcsname\last@language
4165     \wlog{\string\l@#1=\string\language\the\last@language}%
4166     \expandafter\let\csname #1hyphenmins\expandafter\endcsname
4167       \csname\language\name hyphenmins\endcsname
4168     \let\bbl@elt\relax
4169     \edef\bbl@languages{\bbl@languages\bbl@elt{#1}{\the\last@language}{}}%
4170   \fi}
```

`\process@language` The macro `\process@language` is used to process a non-empty line from the ‘configuration file’. It has three arguments, each delimited by white space. The first argument is the ‘name’ of a language; the second is the name of the file that contains the patterns. The optional third argument is the name of a file containing hyphenation exceptions. The first thing to do is call `\addlanguage` to allocate a pattern register and to make that register ‘active’. Then the pattern file is read.

For some hyphenation patterns it is needed to load them with a specific font encoding selected. This can be specified in the file `language.dat` by adding for instance ‘:T1’ to the name of the language. The macro `\bbl@get@enc` extracts the font encoding from the language name and stores it in `\bbl@hyph@enc`. The latter can be used in hyphenation files if you need to set a behavior depending on the given encoding (it is set to empty if no encoding is given).

Pattern files may contain assignments to `\lefthyphenmin` and `\righthyphenmin`. \TeX does not keep track of these assignments. Therefore we try to detect such assignments and store them in the `\<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.

```

4171 \def\process@language#1#2#3{%
4172   \expandafter\addlanguage\csname l@#1\endcsname
4173   \expandafter\language\csname l@#1\endcsname
4174   \edef\language#1{%
4175     \bbl@hook@everylanguage{#1}%
4176     % > luatex
4177     \bbl@get@enc#1::\@@@
4178     \beginngroup
4179       \lefthyphenmin\m@ne
4180       \bbl@hook@loadpatterns{#2}%
4181       % > luatex
4182       \ifnum\lefthyphenmin=\m@ne
4183       \else
4184         \expandafter\xdef\csname #1hyphenmins\endcsname{%
4185           \the\lefthyphenmin\the\righthyphenmin}%
4186       \fi
4187     \endgroup
4188     \def\bbl@tempa{#3}%
4189     \ifx\bbl@tempa\@empty\else
4190       \bbl@hook@loadexceptions{#3}%
4191       % > luatex
4192     \fi
4193     \let\bbl@elt\relax
4194     \edef\bbl@languages{%
4195       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{\bbl@tempa}}%
4196     \ifnum\the\language=\z@
4197       \expandafter\ifx\csname #1hyphenmins\endcsname\relax
4198         \set@hyphenmins\tw@\thr@@\relax
4199       \else
4200         \expandafter\expandafter\expandafter\set@hyphenmins
4201         \csname #1hyphenmins\endcsname
4202       \fi
4203       \the\toks@
4204       \toks@{}%
4205     \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.

```

4206 \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.

```

4207 \def\bbl@hook@everylanguage#1{}
4208 \def\bbl@hook@loadpatterns#1{\input #1\relax}
4209 \let\bbl@hook@loadexceptions\bbl@hook@loadpatterns
4210 \def\bbl@hook@loadkernel#1{%
4211   \def\addlanguage{\csname newlanguage\endcsname}%
4212   \def\adddialect##1##2{%
4213     \global\chardef##1##2\relax

```

```

4214 \wlog{\string##1 = a dialect from \string\language##2}%
4215 \def\iflanguage##1{%
4216 \expandafter\ifx\csname l@##1\endcsname\relax
4217 \nolanerr{##1}%
4218 \else
4219 \ifnum\csname l@##1\endcsname=\language
4220 \expandafter\expandafter\expandafter\@firstoftwo
4221 \else
4222 \expandafter\expandafter\expandafter\@secondoftwo
4223 \fi
4224 \fi}%
4225 \def\providehyphenmins##1##2{%
4226 \expandafter\ifx\csname ##1hyphenmins\endcsname\relax
4227 \namedef{##1hyphenmins}{##2}%
4228 \fi}%
4229 \def\set@hyphenmins##1##2{%
4230 \lefthyphenmin##1\relax
4231 \righthyphenmin##2\relax}%
4232 \def\selectlanguage{%
4233 \errhelp{Selecting a language requires a package supporting it}%
4234 \errmessage{Not loaded}}%
4235 \let\foreignlanguage\selectlanguage
4236 \let\otherlanguage\selectlanguage
4237 \expandafter\let\csname otherlanguage*\endcsname\selectlanguage
4238 \def\bbl@usehooks##1##2{% TODO. Temporary!!
4239 \def\setlocale{%
4240 \errhelp{Find an armchair, sit down and wait}%
4241 \errmessage{Not yet available}}%
4242 \let\uselocale\setlocale
4243 \let\locale\setlocale
4244 \let\selectlocale\setlocale
4245 \let\localename\setlocale
4246 \let\textlocale\setlocale
4247 \let\textlanguage\setlocale
4248 \let\languagetext\setlocale}
4249 \begingroup
4250 \def\AddBabelHook#1#2{%
4251 \expandafter\ifx\csname bbl@hook@#2\endcsname\relax
4252 \def\next{\toks1}%
4253 \else
4254 \def\next{\expandafter\gdef\csname bbl@hook@#2\endcsname####1}%
4255 \fi
4256 \next}
4257 \ifx\directlua\undefined
4258 \ifx\XeTeXinputencoding\undefined\else
4259 \input xebabel.def
4260 \fi
4261 \else
4262 \input luababel.def
4263 \fi
4264 \openin1 = babel-\bbl@format.cfg
4265 \ifeof1
4266 \else
4267 \input babel-\bbl@format.cfg\relax
4268 \fi
4269 \closein1
4270 \endgroup
4271 \bbl@hook@loadkernel{switch.def}

```

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

```
4272 \openin1 = language.dat
```

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

```
4273 \def\languagename{english}%
4274 \ifeof1
4275 \message{I couldn't find the file language.dat,\space
4276         I will try the file hyphen.tex}
4277 \input hyphen.tex\relax
4278 \chardef\l@english\z@
4279 \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`.

```
4280 \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.

```
4281 \loop
4282 \endlinechar\m@ne
4283 \read1 to \bbl@line
4284 \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.

```
4285 \if T\ifeof1F\fi T\relax
4286 \ifx\bbl@line\@empty\else
4287 \edef\bbl@line{\bbl@line\space\space\space}%
4288 \expandafter\process@line\bbl@line\relax
4289 \fi
4290 \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.

```
4291 \begingroup
4292 \def\bbl@elt#1#2#3#4{%
4293 \global\language=#2\relax
4294 \gdef\languagename{#1}%
4295 \def\bbl@elt##1##2##3##4{}}%
4296 \bbl@languages
4297 \endgroup
4298 \fi
4299 \closein1
```

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

```
4300 \if/\the\toks@/\else
4301 \errhelp{language.dat loads no language, only synonyms}
4302 \errmessage{Orphan language synonym}
4303 \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.

```
4304 \let\bbl@line\@undefined
4305 \let\process@line\@undefined
4306 \let\process@synonym\@undefined
4307 \let\process@language\@undefined
4308 \let\bbl@get@enc\@undefined
4309 \let\bbl@hyph@enc\@undefined
4310 \let\bbl@tempa\@undefined
4311 \let\bbl@hook@loadkernel\@undefined
4312 \let\bbl@hook@everylanguage\@undefined
4313 \let\bbl@hook@loadpatterns\@undefined
4314 \let\bbl@hook@loadexceptions\@undefined
4315 \let\patterns\@undefined
```

Here the code for iniT_EX ends.

12 Font handling with fontspec

Add the bidi handler just before luaoftload, 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].

```
4316 <<(*More package options)>> ≡
4317 \chardef\bbl@bidimode\z@
4318 \DeclareOption{bidi=default}{\chardef\bbl@bidimode=\@ne}
4319 \DeclareOption{bidi=basic}{\chardef\bbl@bidimode=101 }
4320 \DeclareOption{bidi=basic-r}{\chardef\bbl@bidimode=102 }
4321 \DeclareOption{bidi=bidi}{\chardef\bbl@bidimode=201 }
4322 \DeclareOption{bidi=bidi-r}{\chardef\bbl@bidimode=202 }
4323 \DeclareOption{bidi=bidi-l}{\chardef\bbl@bidimode=203 }
4324 <</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.

```
4325 <<(*Font selection)>> ≡
4326 \bbl@trace{Font handling with fontspec}
4327 \ifx\ExplSyntaxOn\@undefined\else
4328   \ExplSyntaxOn
4329   \catcode`\ =10
4330   \def\bbl@loadfontspec{%
4331     \usepackage{fontspec}%
4332     \expandafter
4333     \def\csname msg-text->~fontspec/language-not-exist\endcsname##1##2##3##4{%
4334       Font '\l_fontspec_fontname_tl' is using the\\%
4335       default features for language '##1'.\\%
4336       That's usually fine, because many languages\\%
4337       require no specific features, but if the output is\\%
4338       not as expected, consider selecting another font.}
4339     \expandafter
4340     \def\csname msg-text->~fontspec/no-script\endcsname##1##2##3##4{%
```

```

4341     Font '\l_fontspec_fontname_tl' is using the\\%
4342     default features for script '##2'.\\%
4343     That's not always wrong, but if the output is\\%
4344     not as expected, consider selecting another font.}}
4345 \ExplSyntaxOff
4346 \fi
4347 \@onlypreamble\babelfont
4348 \newcommand\babelfont[2][\% 1=langs/scripts 2=fam
4349 \bbl@foreach{#1}{\%
4350 \expandafter\ifx\csname date##1\endcsname\relax
4351 \IfFileExists{babel-##1.tex}%
4352 {\babelprovide{##1}}%
4353 }%
4354 \fi}%
4355 \edef\bbl@tempa{#1}%
4356 \def\bbl@tempb{#2}% Used by \bbl@bblfont
4357 \ifx\fontspec@undefined
4358 \bbl@loadfontspec
4359 \fi
4360 \EnableBabelHook{babel-fontspec}% Just calls \bbl@switchfont
4361 \bbl@bblfont}
4362 \newcommand\bbl@bblfont[2][\% 1=features 2=fontname, @font=rm|sf|tt
4363 \bbl@ifunset{\bbl@tempb family}%
4364 {\bbl@providefam{\bbl@tempb}}%
4365 {\bbl@exp{%
4366 \\\bbl@sreplace\<\bbl@tempb family >%
4367 {\@nameuse{\bbl@tempb default}}{\<\bbl@tempb default>}}}%
4368 % For the default font, just in case:
4369 \bbl@ifunset{\bbl@lsys\language\name}{\bbl@provide@lsys{\language\name}}}%
4370 \expandafter\bbl@ifblank\expandafter{\bbl@tempa}%
4371 {\bbl@csarg\edef{\bbl@tempb dflt@}{<{#1}{#2}}% save bbl@rmdflt@
4372 \bbl@exp{%
4373 \let\<\bbl@\bbl@tempb dflt@\language\name>\<\bbl@\bbl@tempb dflt@>%
4374 \\\bbl@font@set\<\bbl@\bbl@tempb dflt@\language\name>%
4375 \<\bbl@tempb default>\<\bbl@tempb family>}}%
4376 {\bbl@foreach\bbl@tempa{\% ie bbl@rmdflt@lang / *scrt
4377 \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:

```

4378 \def\bbl@providefam#1{\%
4379 \bbl@exp{%
4380 \\\newcommand\<#1default>{\% Just define it
4381 \\\bbl@add@list\\\bbl@font@fams{#1}%
4382 \\\DeclareRobustCommand\<#1family>{\%
4383 \\\not@math@alphabet\<#1family>\relax
4384 \\\fontfamily\<#1default>\selectfont}%
4385 \\\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.

```

4386 \def\bbl@nostdfont#1{\%
4387 \bbl@ifunset{\bbl@WFF@\f@family}%
4388 {\bbl@csarg\gdef{WFF@\f@family}}{\% Flag, to avoid dupl warns
4389 \bbl@infowarn{The current font is not a babel standard family:\\%
4390 #1%
4391 \fontname\font\\%
4392 There is nothing intrinsically wrong with this warning, and\\%
4393 you can ignore it altogether if you do not need these\\%
4394 families. But if they are used in the document, you should be\\%

```

```

4395     aware 'babel' will no set Script and Language for them, so\\%
4396     you may consider defining a new family with \string\babelfont.\\%
4397     See the manual for further details about \string\babelfont.\\%
4398     Reported}}
4399     }}}}
4400 \gdef\bbl@switchfont{%
4401   \bbl@ifunset\bbl@lsys@\language\name\{\bbl@provide@lsys\{\language\name\}\}\%
4402   \bbl@exp{%   eg Arabic -> arabic
4403     \lowercase{\edef\\bbl@tempa{\bbl@cl\{sname\}}}\%
4404   \bbl@foreach\bbl@font@fams{%
4405     \bbl@ifunset\bbl@##1dflt@\language\name\%      (1) language?
4406     {\bbl@ifunset\bbl@##1dflt@*\bbl@tempa\}%      (2) from script?
4407     {\bbl@ifunset\bbl@##1dflt@\}%                  2=F - (3) from generic?
4408     }\%                                             123=F - nothing!
4409     {\bbl@exp{%                                     3=T - from generic
4410       \global\let<\bbl@##1dflt@\language\name>%
4411       \<\bbl@##1dflt@>}}}\%
4412     {\bbl@exp{%                                     2=T - from script
4413       \global\let<\bbl@##1dflt@\language\name>%
4414       \<\bbl@##1dflt@*\bbl@tempa>}}}\%
4415     }}}}                                           1=T - language, already defined
4416 \def\bbl@tempa{\bbl@nostdfont{}}\%
4417 \bbl@foreach\bbl@font@fams{%   don't gather with prev for
4418   \bbl@ifunset\bbl@##1dflt@\language\name\%
4419   {\bbl@cs{famrst@##1}\%
4420     \global\bbl@csarg\let{famrst@##1}\relax}\%
4421   {\bbl@exp{% order is relevant. TODO: but sometimes wrong!
4422     \\bbl@add\\originalTeX{%
4423       \\bbl@font@rst{\bbl@cl{##1dflt}}}\%
4424       \<##1default>\<##1family>{##1}}}\%
4425     \\bbl@font@set<\bbl@##1dflt@\language\name>% the main part!
4426     \<##1default>\<##1family>}}}\%
4427   \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.

```

4428 \ifx\family\undefined\else   % if latex
4429   \ifcase\bbl@engine          % if pdftex
4430     \let\bbl@ckeckstdfonts\relax
4431   \else
4432     \def\bbl@ckeckstdfonts{%
4433       \begingroup
4434       \global\let\bbl@ckeckstdfonts\relax
4435       \let\bbl@tempa\empty
4436       \bbl@foreach\bbl@font@fams{%
4437         \bbl@ifunset\bbl@##1dflt@\}%
4438         {\nameuse{##1family}\%
4439           \bbl@csarg\gdef{WFF@f@family}\}% Flag
4440           \bbl@exp{\\bbl@add\\bbl@tempa{* \<##1family>= \f@family\\}%
4441             \space\space\fontname\font\\}}}\%
4442           \bbl@csarg\xdef{##1dflt@}{f@family}\%
4443           \expandafter\xdef\csname ##1default\endcsname{\f@family}}}\%
4444         }}}}
4445   \ifx\bbl@tempa\empty\else
4446     \bbl@infowarn{The following font families will use the default\\%
4447       settings for all or some languages:\\%
4448       \bbl@tempa
4449       There is nothing intrinsically wrong with it, but\\%
4450       'babel' will no set Script and Language, which could\\%

```

```

4451         be relevant in some languages. If your document uses\\%
4452         these families, consider redefining them with \string\babelfont.\\%
4453         Reported}%
4454         \fi
4455     \endgroup}
4456 \fi
4457 \fi

```

Now the macros defining the font with fontspec.

When there are repeated keys in fontspec, the last value wins. So, we just place the ini settings at the beginning, and user settings will take precedence. We must deactivate temporarily \bbl@mapselect because \selectfont is called internally when a font is defined.

```

4458 \def\bbl@font@set#1#2#3{% eg \bbl@rmdflt@lang \rmdefault \rmfamily
4459 \bbl@xin@{<>}{#1}%
4460 \ifin@
4461 \bbl@exp{\\bbl@fontspec@set\\#1\expandafter\@gobbletwo#1\\#3}%
4462 \fi
4463 \bbl@exp{%          'Unprotected' macros return prev values
4464 \def\\#2{#1}%      eg, \rmdefault{\bbl@rmdflt@lang}
4465 \\bbl@ifsamestring{#2}{\f@family}%
4466 {\\#3%
4467 \\bbl@ifsamestring{\f@series}{\bfdefault}{\\bfseries}{}%
4468 \let\\bbl@tempa\relax}%
4469 {}}}
4470 %      TODO - next should be global?, but even local does its job. I'm
4471 %      still not sure -- must investigate:
4472 \def\bbl@fontspec@set#1#2#3#4{% eg \bbl@rmdflt@lang fnt-opt fnt-nme \xxfamily
4473 \let\bbl@tempe\bbl@mapselect
4474 \let\bbl@mapselect\relax
4475 \let\bbl@temp@fam#4%      eg, '\rmfamily', to be restored below
4476 \let#4\@empty %      Make sure \renewfontfamily is valid
4477 \bbl@exp{%
4478 \let\\bbl@temp@pfam\<\bbl@stripslash#4\space>% eg, '\rmfamily '
4479 \<keys_if_exist:nnF>{fontspec-opentype}{Script/\bbl@cl{sname}}}%
4480 {\\newfontscript{\bbl@cl{sname}}{\bbl@cl{sotf}}}%
4481 \<keys_if_exist:nnF>{fontspec-opentype}{Language/\bbl@cl{lname}}}%
4482 {\\newfontlanguage{\bbl@cl{lname}}{\bbl@cl{lotf}}}%
4483 \\renewfontfamily\\#4%
4484 [\bbl@cs{lsys@\language},#2]{#3}% ie \bbl@exp{..}{#3}
4485 \begingroup
4486 #4%
4487 \xdef#1{\f@family}%      eg, \bbl@rmdflt@lang{FreeSerif(0)}
4488 \endgroup
4489 \let#4\bbl@temp@fam
4490 \bbl@exp{\let\<\bbl@stripslash#4\space>}\bbl@temp@pfam
4491 \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.

```

4492 \def\bbl@font@rst#1#2#3#4{%
4493 \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.

```

4494 \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 :-).

```
4495 \newcommand\babelFSstore[2][{}%
4496   \bbl@ifblank{#1}%
4497     {\bbl@csarg\def{sname@#2}{Latin}}%
4498     {\bbl@csarg\def{sname@#2}{#1}}%
4499   \bbl@provide@dirs{#2}%
4500   \bbl@csarg\ifnum{wdir@#2}>\z@
4501     \let\bbl@beforeforeign\leavevmode
4502     \EnableBabelHook{babel-bidi}%
4503   \fi
4504   \bbl@foreach{#2}{%
4505     \bbl@FSstore{##1}{rm}\rmdefault\bbl@save@rmdefault
4506     \bbl@FSstore{##1}{sf}\sfdefault\bbl@save@sfdefault
4507     \bbl@FSstore{##1}{tt}\ttdefault\bbl@save@ttdefault}}
4508 \def\bbl@FSstore#1#2#3#4{%
4509   \bbl@csarg\edef{#2default#1}{#3}%
4510   \expandafter\addto\csname extras#1\endcsname{%
4511     \let#4#3%
4512     \ifx#3\f@family
4513       \edef#3{\csname bbl@#2default#1\endcsname}%
4514       \fontfamily{#3}\selectfont
4515     \else
4516       \edef#3{\csname bbl@#2default#1\endcsname}%
4517     \fi}%
4518   \expandafter\addto\csname noextras#1\endcsname{%
4519     \ifx#3\f@family
4520       \fontfamily{#4}\selectfont
4521     \fi
4522     \let#3#4}}
4523 \let\bbl@langfeatures\@empty
4524 \def\babelFSfeatures{% make sure \fontspec is redefined once
4525   \let\bbl@ori@fontspec\fontspec
4526   \renewcommand\fontspec[1][{}%
4527     \bbl@ori@fontspec[\bbl@langfeatures##1]}
4528   \let\babelFSfeatures\bbl@FSfeatures
4529   \babelFSfeatures}
4530 \def\bbl@FSfeatures#1#2{%
4531   \expandafter\addto\csname extras#1\endcsname{%
4532     \babel@save\bbl@langfeatures
4533     \edef\bbl@langfeatures{#2,}}
4534 \</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.

```
4535 <<{*Footnote changes}>> ≡
4536 \bbl@trace{Bidi footnotes}
4537 \ifnum\bbl@bidimode>\z@
4538   \def\bbl@footnote#1#2#3{%
4539     \@ifnextchar[%
4540       {\bbl@footnote@o{#1}{#2}{#3}}%
4541       {\bbl@footnote@x{#1}{#2}{#3}}}
4542   \long\def\bbl@footnote@x#1#2#3#4{%
4543     \bgroup
```



```

4544 \select@language@x{\bbl@main@language}%
4545 \bbl@fn@footnote{#2#1{\ignorespaces#4}#3}%
4546 \egroup}
4547 \long\def\bbl@footnote@o#1#2#3[#4]#5{%
4548 \bgroup
4549 \select@language@x{\bbl@main@language}%
4550 \bbl@fn@footnote[#4]{#2#1{\ignorespaces#5}#3}%
4551 \egroup}
4552 \def\bbl@footnotetext#1#2#3{%
4553 \@ifnextchar[%
4554 {\bbl@footnotetext@o{#1}{#2}{#3}}%
4555 {\bbl@footnotetext@x{#1}{#2}{#3}}}
4556 \long\def\bbl@footnotetext@x#1#2#3#4{%
4557 \bgroup
4558 \select@language@x{\bbl@main@language}%
4559 \bbl@fn@footnotetext{#2#1{\ignorespaces#4}#3}%
4560 \egroup}
4561 \long\def\bbl@footnotetext@o#1#2#3[#4]#5{%
4562 \bgroup
4563 \select@language@x{\bbl@main@language}%
4564 \bbl@fn@footnotetext[#4]{#2#1{\ignorespaces#5}#3}%
4565 \egroup}
4566 \def\BabelFootnote#1#2#3#4{%
4567 \ifx\bbl@fn@footnote\undefined
4568 \let\bbl@fn@footnote\footnote
4569 \fi
4570 \ifx\bbl@fn@footnotetext\undefined
4571 \let\bbl@fn@footnotetext\footnotetext
4572 \fi
4573 \bbl@ifblank{#2}%
4574 {\def#1{\bbl@footnote{\@firstofone}{#3}{#4}}
4575 \@namedef{\bbl@stripslash#1text}%
4576 {\bbl@footnotetext{\@firstofone}{#3}{#4}}}%
4577 {\def#1{\bbl@exp{\bbl@footnote{\foreignlanguage{#2}}}{#3}{#4}}%
4578 \@namedef{\bbl@stripslash#1text}%
4579 {\bbl@exp{\bbl@footnotetext{\foreignlanguage{#2}}}{#3}{#4}}}%
4580 \fi
4581 <</Footnote changes>>

```

Now, the code.

```

4582 (*xetex)
4583 \def\BabelStringsDefault{unicode}
4584 \let\xebbl@stop\relax
4585 \AddBabelHook{xetex}{encodedcommands}{%
4586 \def\bbl@tempa{#1}%
4587 \ifx\bbl@tempa\empty
4588 \XeTeXinputencoding"bytes"%
4589 \else
4590 \XeTeXinputencoding"#1"%
4591 \fi
4592 \def\xebbl@stop{\XeTeXinputencoding"utf8"}}
4593 \AddBabelHook{xetex}{stopcommands}{%
4594 \xebbl@stop
4595 \let\xebbl@stop\relax}
4596 \def\bbl@intraspace#1 #2 #3\@@{%
4597 \bbl@csarg\gdef{\xeisp@language}%
4598 {\XeTeXlinebreakskip #1em plus #2em minus #3em\relax}}
4599 \def\bbl@intrapenalty#1\@@{%
4600 \bbl@csarg\gdef{\xeipn@language}%

```

```

4601 {\XeTeXlinebreakpenalty #1\relax}}
4602 \def\bbl@provide@intraspace{%
4603 \bbl@xin@{\bbl@cl{lnbrk}}{s}%
4604 \ifin@else\bbl@xin@{\bbl@cl{lnbrk}}{c}\fi
4605 \ifin@
4606 \bbl@ifunset{\bbl@intsp@{\language\language}}{s}%
4607 {\expandafter\ifx\csname\bbl@intsp@{\language\language}\endcsname\empty\else
4608 \ifx\bbl@KVP@intraspace\@nil
4609 \bbl@exp{%
4610 \bbl@intraspace\bbl@cl{intsp}\@{}%
4611 \fi
4612 \ifx\bbl@KVP@intrapenalty\@nil
4613 \bbl@intrapenalty0\@{}
4614 \fi
4615 \fi
4616 \ifx\bbl@KVP@intraspace\@nil\else % We may override the ini
4617 \expandafter\bbl@intraspace\bbl@KVP@intraspace\@{}
4618 \fi
4619 \ifx\bbl@KVP@intrapenalty\@nil\else
4620 \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@{}
4621 \fi
4622 \bbl@exp{%
4623 \bbl@add\<extras\language>{%
4624 \XeTeXlinebreaklocale "\bbl@cl{tbcpr}"%
4625 \<bbl@xeisp@\language>%
4626 \<bbl@xeipn@\language>%
4627 \bbl@toglobal\<extras\language>%
4628 \bbl@add\<noextras\language>{%
4629 \XeTeXlinebreaklocale "en"%
4630 \bbl@toglobal\<noextras\language>%
4631 \ifx\bbl@ispace\@undefined
4632 \gdef\bbl@ispace{\bbl@cl{xeisp}}%
4633 \ifx\AtBeginDocument\@notprerr
4634 \expandafter\@secondoftwo % to execute right now
4635 \fi
4636 \AtBeginDocument{%
4637 \expandafter\bbl@add
4638 \csname selectfont \endcsname{\bbl@ispace}%
4639 \expandafter\bbl@toglobal\csname selectfont \endcsname}%
4640 \fi}%
4641 \fi}
4642 \ifx\DisableBabelHook\@undefined\endinput\fi
4643 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
4644 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@ckeckstdfonts}
4645 \DisableBabelHook{babel-fontspec}
4646 <<Font selection>>
4647 \input txtbabel.def
4648 </xetex>

```

13.2 Layout

In progress.

Note elements like headlines and margins can be modified easily with packages like fancyhdr, typearea or titlesp, and geometry.

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

Consider txtbabel as a shorthand for *tex-xet babel*, which is the bidi model in both pdftex

and xetex.

```
4649 (*texxet)
4650 \providecommand\bbl@provide@intraspace{}
4651 \bbl@trace{Redefinitions for bidi layout}
4652 \def\bbl@sspre@caption{%
4653   \bbl@exp{\everyhbox{\bbl@textdir\bbl@cs{wdir\bbl@main@language}}}}
4654 \ifx\bbl@opt@layout\@nnil\endinput\fi % No layout
4655 \def\bbl@startskip{\ifcase\bbl@thepardir\leftskip\else\rightskip\fi}
4656 \def\bbl@endskip{\ifcase\bbl@thepardir\rightskip\else\leftskip\fi}
4657 \ifx\bbl@beforeforeign\leavevmode % A poor test for bidi=
4658   \def\hangfrom#1{%
4659     \setbox\@tempboxa\hbox{#1}%
4660     \hangindent\ifcase\bbl@thepardir\wd\@tempboxa\else-\wd\@tempboxa\fi
4661     \noindent\box\@tempboxa}
4662 \def\raggedright{%
4663   \let\@centercr
4664   \bbl@startskip\z@skip
4665   \@rightskip\@flushglue
4666   \bbl@endskip\@rightskip
4667   \parindent\z@
4668   \parfillskip\bbl@startskip}
4669 \def\raggedleft{%
4670   \let\@centercr
4671   \bbl@startskip\@flushglue
4672   \bbl@endskip\z@skip
4673   \parindent\z@
4674   \parfillskip\bbl@endskip}
4675 \fi
4676 \IfBabelLayout{lists}
4677   {\bbl@sreplace\list
4678     {\@totalleftmargin\leftmargin}{\@totalleftmargin\bbl@listleftmargin}%
4679     \def\bbl@listleftmargin{%
4680       \ifcase\bbl@thepardir\leftmargin\else\rightmargin\fi}%
4681     \ifcase\bbl@engine
4682       \def\labelenumii{}\theenumii{}\pdfTeX doesn't reverse ()
4683       \def\p@enumiii{\p@enumii}\theenumii{}
4684     \fi
4685     \bbl@sreplace\@verbatim
4686     {\leftskip\@totalleftmargin}%
4687     {\bbl@startskip\textwidth
4688       \advance\bbl@startskip-\linewidth}%
4689     \bbl@sreplace\@verbatim
4690     {\rightskip\z@skip}%
4691     {\bbl@endskip\z@skip}}%
4692   {}
4693 \IfBabelLayout{contents}
4694   {\bbl@sreplace\@dottedtocline{\leftskip}{\bbl@startskip}%
4695     \bbl@sreplace\@dottedtocline{\rightskip}{\bbl@endskip}}
4696   {}
4697 \IfBabelLayout{columns}
4698   {\bbl@sreplace\@outputdblcol{\hb@xt@\textwidth}{\bbl@outputbox}%
4699     \def\bbl@outputbox#1{%
4700       \hb@xt@\textwidth{%
4701         \hskip\columnwidth
4702         \hfil
4703         {\normalcolor\vrule \@width\columnseprule}%
4704         \hfil
4705         \hb@xt@\columnwidth{\box\@leftcolumn \hss}%
4706       }
4707     }
4708   }
```

```

4706      \hskip-\textwidth
4707      \hb@xt@\columnwidth{\box\@outputbox \hss}%
4708      \hskip\columnsep
4709      \hskip\columnwidth}}}%
4710  {}
4711  <<Footnote changes>>
4712  \IfBabelLayout{footnotes}%
4713  {\BabelFootnote\footnote\language\language{}{}}%
4714  \BabelFootnote\localfootnote\language\language{}{}}%
4715  \BabelFootnote\mainfootnote{}{}}{}
4716  {}

```

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.

```

4717 \IfBabelLayout{counters}%
4718 {\let\bbbl@latinarabic=\@arabic
4719  \def\@arabic#1{\babelsublr{\bbbl@latinarabic#1}}}%
4720 \let\bbbl@asciroman=\@roman
4721 \def\@roman#1{\babelsublr{\ensureascii{\bbbl@asciroman#1}}}%
4722 \let\bbbl@asciiRoman=\@Roman
4723 \def\@Roman#1{\babelsublr{\ensureascii{\bbbl@asciiRoman#1}}}}{}
4724 </texxet>

```

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, `\bbbl@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 files 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).

```

4725 (*luatex)
4726 \ifx\AddBabelHook\undefined % When plain.def, babel.sty starts
4727 \bbl@trace{Read language.dat}
4728 \ifx\bbl@readstream\undefined
4729 \csname newread\endcsname\bbl@readstream
4730 \fi
4731 \begingroup
4732 \toks@{}
4733 \count@ \z@ % 0=start, 1=0th, 2=normal
4734 \def\bbl@process@line#1#2 #3 #4 {%
4735   \ifx=#1%
4736     \bbl@process@synonym{#2}%
4737   \else
4738     \bbl@process@language{#1#2}{#3}{#4}%
4739   \fi
4740   \ignorespaces}
4741 \def\bbl@manylang{%
4742   \ifnum\bbl@last>\@ne
4743     \bbl@info{Non-standard hyphenation setup}%
4744   \fi
4745   \let\bbl@manylang\relax}
4746 \def\bbl@process@language#1#2#3{%
4747   \ifcase\count@
4748     \@ifundefined{zth#1}{\count@\tw@}{\count@\@ne}%
4749   \or
4750     \count@\tw@
4751   \fi
4752   \ifnum\count@=\tw@
4753     \expandafter\addlanguage\csname l@#1\endcsname
4754     \language\allocationnumber
4755     \chardef\bbl@last\allocationnumber
4756     \bbl@manylang
4757     \let\bbl@elt\relax
4758     \xdef\bbl@languages{%
4759       \bbl@languages\bbl@elt{#1}{\the\language}{#2}{#3}}%
4760   \fi
4761   \the\toks@
4762   \toks@{}}
4763 \def\bbl@process@synonym@aux#1#2{%
4764   \global\expandafter\chardef\csname l@#1\endcsname#2\relax
4765   \let\bbl@elt\relax
4766   \xdef\bbl@languages{%
4767     \bbl@languages\bbl@elt{#1}{#2}{}}}%
4768 \def\bbl@process@synonym#1{%
4769   \ifcase\count@
4770     \toks@\expandafter{\the\toks@\relax\bbl@process@synonym{#1}}%
4771   \or
4772     \@ifundefined{zth#1}{\bbl@process@synonym@aux{#1}{0}}{}%
4773   \else
4774     \bbl@process@synonym@aux{#1}{\the\bbl@last}%
4775   \fi}
4776 \ifx\bbl@languages\undefined % Just a (sensible?) guess
4777   \chardef\l@english\z@
4778   \chardef\l@USenglish\z@

```

```

4779 \chardef\bbl@last\z@
4780 \global\@namedef{bbl@hyphendata@0}{\hyphen.tex}{}
4781 \gdef\bbl@languages{%
4782   \bbl@elt{english}{0}{\hyphen.tex}{}%
4783   \bbl@elt{USenglish}{0}{}{}}
4784 \else
4785   \global\let\bbl@languages@format\bbl@languages
4786   \def\bbl@elt#1#2#3#4{% Remove all except language 0
4787     \ifnum#2>\z@\else
4788       \noexpand\bbl@elt{#1}{#2}{#3}{#4}%
4789     \fi}%
4790   \xdef\bbl@languages{\bbl@languages}%
4791 \fi
4792 \def\bbl@elt#1#2#3#4{\@namedef{zth@#1}{} } % Define flags
4793 \bbl@languages
4794 \openin\bbl@readstream=language.dat
4795 \ifeof\bbl@readstream
4796   \bbl@warning{I couldn't find language.dat. No additional\\%
4797     patterns loaded. Reported}%
4798 \else
4799   \loop
4800     \endlinechar\m@ne
4801     \read\bbl@readstream to \bbl@line
4802     \endlinechar`\^^M
4803     \if T\ifeof\bbl@readstream F\fi T\relax
4804     \ifx\bbl@line\empty\else
4805       \edef\bbl@line{\bbl@line\space\space\space}%
4806       \expandafter\bbl@process@line\bbl@line\relax
4807     \fi
4808   \repeat
4809 \fi
4810 \endgroup
4811 \bbl@trace{Macros for reading patterns files}
4812 \def\bbl@get@enc#1:#2:#3\@@@{\def\bbl@hyph@enc{#2}}
4813 \ifx\babelcatcodetablenum\undefined
4814   \ifx\newcatcodetable\undefined
4815     \def\babelcatcodetablenum{5211}
4816     \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4817   \else
4818     \newcatcodetable\babelcatcodetablenum
4819     \newcatcodetable\bbl@pattcodes
4820   \fi
4821 \else
4822   \def\bbl@pattcodes{\numexpr\babelcatcodetablenum+1\relax}
4823 \fi
4824 \def\bbl@luapatterns#1#2{%
4825   \bbl@get@enc#1::\@@@
4826   \setbox\z@\hbox\bgroup
4827   \begingroup
4828     \savecatcodetable\babelcatcodetablenum\relax
4829     \initcatcodetable\bbl@pattcodes\relax
4830     \catcodetable\bbl@pattcodes\relax
4831     \catcode`\#=6 \catcode`\$=3 \catcode`\&=4 \catcode`\^=7
4832     \catcode`\_ =8 \catcode`\{=1 \catcode`\}=2 \catcode`\~=13
4833     \catcode`\@=11 \catcode`\^^I=10 \catcode`\^^J=12
4834     \catcode`\<=12 \catcode`\>=12 \catcode`\*=12 \catcode`\.=12
4835     \catcode`\-=12 \catcode`\/=12 \catcode`\[=12 \catcode`\]=12
4836     \catcode`\'=12 \catcode`\'=12 \catcode`\`=12
4837     \input #1\relax

```

```

4838     \catcodetable\babelcatcodetablenum\relax
4839 \endgroup
4840 \def\bbl@tempa{#2}%
4841 \ifx\bbl@tempa\@empty\else
4842     \input #2\relax
4843 \fi
4844 \egroup}%
4845 \def\bbl@patterns@lua#1{%
4846 \language=\expandafter\ifx\csname l@#1:\f@encoding\endcsname\relax
4847     \csname l@#1\endcsname
4848     \edef\bbl@tempa{#1}%
4849 \else
4850     \csname l@#1:\f@encoding\endcsname
4851     \edef\bbl@tempa{#1:\f@encoding}%
4852 \fi\relax
4853 \@namedef{lu@texhyphen@loaded@the\language}{}% Temp
4854 \@ifundefined{bbl@hyphendata@the\language}%
4855     {\def\bbl@elt##1##2##3##4{%
4856         \ifnum##2=\csname l@bbl@tempa\endcsname % #2=spanish, dutch:OT1...
4857         \def\bbl@tempb{##3}%
4858         \ifx\bbl@tempb\@empty\else % if not a synonymous
4859             \def\bbl@tempc{##3}{##4}%
4860         \fi
4861         \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4862         \fi}%
4863     \bbl@languages
4864     \@ifundefined{bbl@hyphendata@the\language}%
4865         {\bbl@info{No hyphenation patterns were set for\%
4866             language '\bbl@tempa'. Reported}}%
4867         {\expandafter\expandafter\expandafter\bbl@luapatterns
4868         \csname bbl@hyphendata@the\language\endcsname}}}%
4869 \endinput\fi
4870 % Here ends \ifx\AddBabelHook\@undefined
4871 % A few lines are only read by hyphen.cfg
4872 \ifx\DisableBabelHook\@undefined
4873     \AddBabelHook{luatex}{everylanguage}{%
4874         \def\process@language##1##2##3{%
4875             \def\process@line####1####2 ####3 ####4 {}}}
4876     \AddBabelHook{luatex}{loadpatterns}{%
4877         \input #1\relax
4878         \expandafter\gdef\csname bbl@hyphendata@the\language\endcsname
4879             {{#1}}}}
4880     \AddBabelHook{luatex}{loadexceptions}{%
4881         \input #1\relax
4882         \def\bbl@tempb##1##2{{##1}{##2}}%
4883         \expandafter\xdef\csname bbl@hyphendata@the\language\endcsname
4884             {\expandafter\expandafter\expandafter\bbl@tempb
4885             \csname bbl@hyphendata@the\language\endcsname}}
4886 \endinput\fi
4887 % Here stops reading code for hyphen.cfg
4888 % The following is read the 2nd time it's loaded
4889 \begingroup % TODO - to a lua file
4890 \catcode`\%=12
4891 \catcode`\'=12
4892 \catcode`\%=12
4893 \catcode`\:=12
4894 \directlua{
4895     Babel = Babel or {}
4896     function Babel.bytes(line)

```

```

4897     return line:gsub("(.)",
4898         function (chr) return unicode.utf8.char(string.byte(chr)) end)
4899 end
4900 function Babel.begin_process_input()
4901     if luatexbase and luatexbase.add_to_callback then
4902         luatexbase.add_to_callback('process_input_buffer',
4903             Babel.bytes,'Babel.bytes')
4904     else
4905         Babel.callback = callback.find('process_input_buffer')
4906         callback.register('process_input_buffer',Babel.bytes)
4907     end
4908 end
4909 function Babel.end_process_input ()
4910     if luatexbase and luatexbase.remove_from_callback then
4911         luatexbase.remove_from_callback('process_input_buffer','Babel.bytes')
4912     else
4913         callback.register('process_input_buffer',Babel.callback)
4914     end
4915 end
4916 function Babel.addpatterns(pp, lg)
4917     local lg = lang.new(lg)
4918     local pats = lang.patterns(lg) or ''
4919     lang.clear_patterns(lg)
4920     for p in pp:gmatch('[^%s]+') do
4921         ss = ''
4922         for i in string.utfcharacters(p:gsub('%d', '')) do
4923             ss = ss .. '%d?' .. i
4924         end
4925         ss = ss:gsub('^%%d%?%.','%%%.') .. '%d?'
4926         ss = ss:gsub('%.%%d%?$','%%%.')
4927         pats, n = pats:gsub('%s' .. ss .. '%s', ' ' .. p .. ' ')
4928         if n == 0 then
4929             tex.sprint(
4930                 [[\string\csname\space bbl@info\endcsname{New pattern: }]]
4931                 .. p .. [[]])
4932             pats = pats .. ' ' .. p
4933         else
4934             tex.sprint(
4935                 [[\string\csname\space bbl@info\endcsname{Renew pattern: }]]
4936                 .. p .. [[]])
4937         end
4938     end
4939     lang.patterns(lg, pats)
4940 end
4941 }
4942 \endgroup
4943 \ifx\newattribute\@undefined\else
4944     \newattribute\bbl@attr@locale
4945     \directlua{ Babel.attr_locale = luatexbase.registernumber'bbl@attr@locale'}
4946     \AddBabelHook{luatex}{beforeextras}{%
4947         \setattribute\bbl@attr@locale\localeid}
4948 \fi
4949 \def\BabelStringsDefault{unicode}
4950 \let\luabbl@stop\relax
4951 \AddBabelHook{luatex}{encodedcommands}{%
4952     \def\bbl@tempa{utf8}\def\bbl@tempb{#1}%
4953     \ifx\bbl@tempa\bbl@tempb\else
4954         \directlua{Babel.begin_process_input()}%
4955     \def\luabbl@stop{%

```



```

4956 \directlua{Babel.end_process_input()}}%
4957 \fi}%
4958 \AddBabelHook{luatex}{stopcommands}{%
4959 \luabbbl@stop
4960 \let\luabbbl@stop\relax}
4961 \AddBabelHook{luatex}{patterns}{%
4962 \@ifundefined{bbl@hyphendata@the\language}%
4963 {\def\bbl@elt##1##2##3##4{%
4964 \ifnum##2=\csname l@##2\endcsname % #2=spanish, dutch:OT1...
4965 \def\bbl@tempb{##3}%
4966 \ifx\bbl@tempb\empty\else % if not a synonymous
4967 \def\bbl@tempc{##3}{##4}}%
4968 \fi
4969 \bbl@csarg\xdef{hyphendata@##2}{\bbl@tempc}%
4970 \fi}%
4971 \bbl@languages
4972 \@ifundefined{bbl@hyphendata@the\language}%
4973 {\bbl@info{No hyphenation patterns were set for\%
4974 language '#2'. Reported}}%
4975 {\expandafter\expandafter\expandafter\bbl@luapatterns
4976 \csname bbl@hyphendata@the\language\endcsname}}}%
4977 \@ifundefined{bbl@patterns@}{}%
4978 \begingroup
4979 \bbl@xin@{\number\language,}{\bbl@pttnlist}%
4980 \ifin@else
4981 \ifx\bbl@patterns@\empty\else
4982 \directlua{ Babel.addpatterns(
4983 [[\bbl@patterns@]], \number\language) }%
4984 \fi
4985 \@ifundefined{bbl@patterns@#1}%
4986 \empty
4987 {\directlua{ Babel.addpatterns(
4988 [[\space\csname bbl@patterns@#1\endcsname]],
4989 \number\language) }}%
4990 \xdef\bbl@pttnlist{\bbl@pttnlist\number\language,}%
4991 \fi
4992 \endgroup}%
4993 \bbl@exp{%
4994 \bbl@ifunset{bbl@prehc@\languagename}{}%
4995 {\bbl@ifblank{\bbl@cs{prehc@\languagename}}{}}%
4996 {\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.

```

4997 \@onlypreamble\babelpatterns
4998 \AtEndOfPackage{%
4999 \newcommand\babelpatterns[2][\empty]{%
5000 \ifx\bbl@patterns@\relax
5001 \let\bbl@patterns@\empty
5002 \fi
5003 \ifx\bbl@pttnlist\empty\else
5004 \bbl@warning{%
5005 You must not intermingle \string\selectlanguage\space and\%
5006 \string\babelpatterns\space or some patterns will not\%
5007 be taken into account. Reported}%
5008 \fi
5009 \ifx\@empty#1%

```

```

5010     \protected@edef\bbl@patterns@{\bbl@patterns@\space#2}%
5011   \else
5012     \edef\bbl@tempb{\zap@space#1 \@empty}%
5013     \bbl@for\bbl@tempa\bbl@tempb{%
5014       \bbl@fixname\bbl@tempa
5015       \bbl@iflanguage\bbl@tempa{%
5016         \bbl@csarg\protected@edef{patterns@\bbl@tempa}{%
5017           \ifundefined\bbl@patterns@\bbl@tempa}%
5018           \@empty
5019           {\csname bbl@patterns@\bbl@tempa\endcsname\space}%
5020           #2}}}%
5021   \fi}}

```

13.4 Southeast Asian scripts

First, some general code for line breaking, used by `\babelposthyphenation`. *In progress*. Replace regular (ie, implicit) discretionary by spaceskips, based on the previous glyph (which I think makes sense, because the hyphen and the previous char go always together). Other discretionary are not touched. See Unicode UAX 14.

```

5022 % TODO - to a lua file
5023 \directlua{
5024   Babel = Babel or {}
5025   Babel.linebreaking = Babel.linebreaking or {}
5026   Babel.linebreaking.before = {}
5027   Babel.linebreaking.after = {}
5028   Babel.locale = {} % Free to use, indexed with \localeid
5029   function Babel.linebreaking.add_before(func)
5030     tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5031     table.insert(Babel.linebreaking.before , func)
5032   end
5033   function Babel.linebreaking.add_after(func)
5034     tex.print([[noexpand\csname bbl@luahyphenate\endcsname]])
5035     table.insert(Babel.linebreaking.after, func)
5036   end
5037 }
5038 \def\bbl@intraspace#1 #2 #3\@@{%
5039   \directlua{
5040     Babel = Babel or {}
5041     Babel.intraspaces = Babel.intraspaces or {}
5042     Babel.intraspaces['\csname bbl@sbcpr@\languagename\endcsname'] = %
5043       {b = #1, p = #2, m = #3}
5044     Babel.locale_props[\the\localeid].intraspace = %
5045       {b = #1, p = #2, m = #3}
5046   }}
5047 \def\bbl@intrapenalty#1\@@{%
5048   \directlua{
5049     Babel = Babel or {}
5050     Babel.intrapenalties = Babel.intrapenalties or {}
5051     Babel.intrapenalties['\csname bbl@sbcpr@\languagename\endcsname'] = #1
5052     Babel.locale_props[\the\localeid].intrapenalty = #1
5053   }}
5054 \begingroup
5055 \catcode`\%=12
5056 \catcode`\^=14
5057 \catcode`\'=12
5058 \catcode`\~=12
5059 \gdef\bbl@seaintraspace{^
5060   \let\bbl@seaintraspace\relax

```

```

5061 \directlua{
5062   Babel = Babel or {}
5063   Babel.sea_enabled = true
5064   Babel.sea_ranges = Babel.sea_ranges or {}
5065   function Babel.set_chranges (script, chrng)
5066     local c = 0
5067     for s, e in string.gmatch(chrng..' ', '(.-%.%.(-)%s') do
5068       Babel.sea_ranges[script..c]={tonumber(s,16), tonumber(e,16)}
5069       c = c + 1
5070     end
5071   end
5072   function Babel.sea_disc_to_space (head)
5073     local sea_ranges = Babel.sea_ranges
5074     local last_char = nil
5075     local quad = 655360      ^% 10 pt = 655360 = 10 * 65536
5076     for item in node.traverse(head) do
5077       local i = item.id
5078       if i == node.id'glyph' then
5079         last_char = item
5080       elseif i == 7 and item.subtype == 3 and last_char
5081         and last_char.char > 0x0C99 then
5082         quad = font.getfont(last_char.font).size
5083         for lg, rg in pairs(sea_ranges) do
5084           if last_char.char > rg[1] and last_char.char < rg[2] then
5085             lg = lg:sub(1, 4)  ^% Remove trailing number of, eg, Cyril1
5086             local intraspace = Babel.intraspaces[lg]
5087             local intrapenalty = Babel.intrapenalties[lg]
5088             local n
5089             if intrapenalty ~= 0 then
5090               n = node.new(14, 0)      ^% penalty
5091               n.penalty = intrapenalty
5092               node.insert_before(head, item, n)
5093             end
5094             n = node.new(12, 13)      ^% (glue, spaceskip)
5095             node.setglue(n, intraspace.b * quad,
5096               intraspace.p * quad,
5097               intraspace.m * quad)
5098             node.insert_before(head, item, n)
5099             node.remove(head, item)
5100           end
5101         end
5102       end
5103     end
5104   end
5105 }^^
5106 \bbl@luahyphenate}
5107 \catcode`\%=14
5108 \gdef\bbl@cjkintraspaces{%
5109   \let\bbl@cjkintraspaces\relax
5110   \directlua{
5111     Babel = Babel or {}
5112     require'babel-data-cjk.lua'
5113     Babel.cjk_enabled = true
5114     function Babel.cjk_linebreak(head)
5115       local GLYPH = node.id'glyph'
5116       local last_char = nil
5117       local quad = 655360      % 10 pt = 655360 = 10 * 65536
5118       local last_class = nil
5119       local last_lang = nil

```

```

5120
5121     for item in node.traverse(head) do
5122         if item.id == GLYPH then
5123
5124             local lang = item.lang
5125
5126             local LOCALE = node.get_attribute(item,
5127                 luatexbase.registernumber'bbl@attr@locale')
5128             local props = Babel.locale_props[LOCALE]
5129
5130             local class = Babel.cjk_class[item.char].c
5131
5132             if class == 'cp' then class = 'cl' end % ]) as CL
5133             if class == 'id' then class = 'I' end
5134
5135             local br = 0
5136             if class and last_class and Babel.cjk_breaks[last_class][class] then
5137                 br = Babel.cjk_breaks[last_class][class]
5138             end
5139
5140             if br == 1 and props.linebreak == 'c' and
5141                 lang ~= \the\l@nohyphenation\space and
5142                 last_lang ~= \the\l@nohyphenation then
5143                 local intrapenalty = props.intrapenalty
5144                 if intrapenalty ~= 0 then
5145                     local n = node.new(14, 0)    % penalty
5146                     n.penalty = intrapenalty
5147                     node.insert_before(head, item, n)
5148                 end
5149                 local intraspace = props.intraspace
5150                 local n = node.new(12, 13)    % (glue, spaceskip)
5151                 node.setglue(n, intraspace.b * quad,
5152                     intraspace.p * quad,
5153                     intraspace.m * quad)
5154                 node.insert_before(head, item, n)
5155             end
5156
5157             if font.getfont(item.font) then
5158                 quad = font.getfont(item.font).size
5159             end
5160             last_class = class
5161             last_lang = lang
5162         else % if penalty, glue or anything else
5163             last_class = nil
5164         end
5165     end
5166     lang.hyphenate(head)
5167 end
5168 }%
5169 \bbl@luahyphenate}
5170 \gdef\bbl@luahyphenate{%
5171     \let\bbl@luahyphenate\relax
5172     \directlua{
5173         luatexbase.add_to_callback('hyphenate',
5174             function (head, tail)
5175                 if Babel.linebreaking.before then
5176                     for k, func in ipairs(Babel.linebreaking.before) do
5177                         func(head)
5178                     end

```

```

5179 end
5180 if Babel.cjk_enabled then
5181     Babel.cjk_linebreak(head)
5182 end
5183 lang.hyphenate(head)
5184 if Babel.linebreaking.after then
5185     for k, func in ipairs(Babel.linebreaking.after) do
5186         func(head)
5187     end
5188 end
5189 if Babel.sea_enabled then
5190     Babel.sea_disc_to_space(head)
5191 end
5192 end,
5193 'Babel.hyphenate')
5194 }
5195 }
5196 \endgroup
5197 \def\bbl@provide@intraspace{%
5198     \bbl@ifunset{\bbl@intsp@\language\endcsname\empty\else
5199         {\expandafter\ifx\csname\bbl@intsp@\language\endcsname\empty\else
5200             \bbl@xin{\bbl@cl{lncrk}}{c}%
5201             \lfin@% cjk
5202             \bbl@cjkintraspace
5203             \directlua{
5204                 Babel = Babel or {}
5205                 Babel.locale_props = Babel.locale_props or {}
5206                 Babel.locale_props[\the\localeid].linebreak = 'c'
5207             }%
5208             \bbl@exp{\bbl@intraspace\bbl@cl{intsp}}{\bbl@}%
5209             \ifx\bbl@KVP@intrapenalty\@nil
5210                 \bbl@intrapenalty0\@@
5211             \fi
5212         \else % sea
5213             \bbl@seaintraspace
5214             \bbl@exp{\bbl@intraspace\bbl@cl{intsp}}{\bbl@}%
5215             \directlua{
5216                 Babel = Babel or {}
5217                 Babel.sea_ranges = Babel.sea_ranges or {}
5218                 Babel.set_chranges('\bbl@cl{sbcpr}',
5219                                     '\bbl@cl{chrng}')
5220             }%
5221             \ifx\bbl@KVP@intrapenalty\@nil
5222                 \bbl@intrapenalty0\@@
5223             \fi
5224         \fi
5225     \fi
5226     \ifx\bbl@KVP@intrapenalty\@nil\else
5227         \expandafter\bbl@intrapenalty\bbl@KVP@intrapenalty\@@
5228     \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.

```
5229 \AddBabelHook{babel-fontspec}{afterextras}{\bbl@switchfont}
5230 \AddBabelHook{babel-fontspec}{beforestart}{\bbl@cckstfont}
5231 \DisableBabelHook{babel-fontspec}
5232 <<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.

```
5233% TODO - to a lua file
5234 \directlua{
5235 Babel.script_blocks = {
5236   ['Arab'] = {{0x0600, 0x06FF}, {0x08A0, 0x08FF}, {0x0750, 0x077F},
5237             {0xFE70, 0xFEFF}, {0xFB50, 0xFDFF}, {0x1EE00, 0x1EEFF}},
5238   ['Armn'] = {{0x0530, 0x058F}},
5239   ['Beng'] = {{0x0980, 0x09FF}},
5240   ['Cher'] = {{0x13A0, 0x13FF}, {0xAB70, 0xABBF}},
5241   ['Copt'] = {{0x03E2, 0x03EF}, {0x2C80, 0x2CFF}, {0x102E0, 0x102FF}},
5242   ['Cyr1'] = {{0x0400, 0x04FF}, {0x0500, 0x052F}, {0x1C80, 0x1C8F},
5243             {0x2DE0, 0x2DFF}, {0xA640, 0xA69F}},
5244   ['Deva'] = {{0x0900, 0x097F}, {0xA8E0, 0xA8FF}},
5245   ['Ethi'] = {{0x1200, 0x137F}, {0x1380, 0x139F}, {0x2D80, 0x2DDF},
5246             {0xAB00, 0xAB2F}},
5247   ['Geor'] = {{0x10A0, 0x10FF}, {0x2D00, 0x2D2F}},
5248   % Don't follow strictly Unicode, which places some Coptic letters in
5249   % the 'Greek and Coptic' block
5250   ['Grek'] = {{0x0370, 0x03E1}, {0x03F0, 0x03FF}, {0x1F00, 0x1FFF}},
5251   ['Hans'] = {{0x2E80, 0x2EFF}, {0x3000, 0x303F}, {0x31C0, 0x31EF},
5252             {0x3300, 0x33FF}, {0x3400, 0x4DBF}, {0x4E00, 0x9FFF},
5253             {0xF900, 0xFAFF}, {0xFE30, 0xFE4F}, {0xFF00, 0xFFEF},
5254             {0x20000, 0x2A6DF}, {0x2A700, 0x2B73F},
5255             {0x2B740, 0x2B81F}, {0x2B820, 0x2CEAF},
5256             {0x2CEB0, 0x2EBEF}, {0x2F800, 0x2FA1F}},
5257   ['Hebr'] = {{0x0590, 0x05FF}},
5258   ['Jpan'] = {{0x3000, 0x303F}, {0x3040, 0x309F}, {0x30A0, 0x30FF},
5259             {0x4E00, 0x9FAF}, {0xFF00, 0xFFEF}},
5260   ['Khmr'] = {{0x1780, 0x17FF}, {0x19E0, 0x19FF}},
5261   ['Knda'] = {{0x0C80, 0x0CFF}},
5262   ['Kore'] = {{0x1100, 0x11FF}, {0x3000, 0x303F}, {0x3130, 0x318F},
5263             {0x4E00, 0x9FAF}, {0xA960, 0xA97F}, {0xAC00, 0xD7AF},
5264             {0xD7B0, 0xD7FF}, {0xFF00, 0xFFEF}},
5265   ['Lao'] = {{0x0E80, 0x0EFF}},
5266   ['Latn'] = {{0x0000, 0x007F}, {0x0080, 0x00FF}, {0x0100, 0x017F},
5267             {0x0180, 0x024F}, {0x1E00, 0x1EFF}, {0x2C60, 0x2C7F},
5268             {0xA720, 0xA7FF}, {0xAB30, 0xAB6F}},
5269   ['Mahj'] = {{0x11150, 0x1117F}},
5270   ['Mlym'] = {{0x0D00, 0x0D7F}},
```

```

5271 ['Mymr'] = {{0x1000, 0x109F}, {0xAA60, 0xAA7F}, {0xA9E0, 0xA9FF}},
5272 ['Orya'] = {{0x0B00, 0x0B7F}},
5273 ['Sinh'] = {{0x0D80, 0x0DFF}, {0x111E0, 0x111FF}},
5274 ['Sycr'] = {{0x0700, 0x074F}, {0x0860, 0x086F}},
5275 ['Taml'] = {{0x0B80, 0x0BFF}},
5276 ['Telu'] = {{0x0C00, 0x0C7F}},
5277 ['Tfng'] = {{0x2D30, 0x2D7F}},
5278 ['Thai'] = {{0x0E00, 0x0E7F}},
5279 ['Tibt'] = {{0x0F00, 0x0FFF}},
5280 ['Vaii'] = {{0xA500, 0xA63F}},
5281 ['Yiii'] = {{0xA000, 0xA48F}, {0xA490, 0xA4CF}}
5282 }
5283
5284 Babel.script_blocks.Cyrs = Babel.script_blocks.Cyrl
5285 Babel.script_blocks.Hant = Babel.script_blocks.Hans
5286 Babel.script_blocks.Kana = Babel.script_blocks.Jpan
5287
5288 function Babel.locale_map(head)
5289   if not Babel.locale_mapped then return head end
5290
5291   local LOCALE = luatexbase.registernumber'bbl@attr@locale'
5292   local GLYPH = node.id('glyph')
5293   local inmath = false
5294   local toloc_save
5295   for item in node.traverse(head) do
5296     local toloc
5297     if not inmath and item.id == GLYPH then
5298       % Optimization: build a table with the chars found
5299       if Babel.chr_to_loc[item.char] then
5300         toloc = Babel.chr_to_loc[item.char]
5301       else
5302         for lc, maps in pairs(Babel.loc_to_scr) do
5303           for _, rg in pairs(maps) do
5304             if item.char >= rg[1] and item.char <= rg[2] then
5305               Babel.chr_to_loc[item.char] = lc
5306               toloc = lc
5307               break
5308             end
5309           end
5310         end
5311       end
5312       % Now, take action, but treat composite chars in a different
5313       % fashion, because they 'inherit' the previous locale. Not yet
5314       % optimized.
5315       if not toloc and
5316         (item.char >= 0x0300 and item.char <= 0x036F) or
5317         (item.char >= 0x1AB0 and item.char <= 0x1AFF) or
5318         (item.char >= 0x1DC0 and item.char <= 0x1DFF) then
5319         toloc = toloc_save
5320       end
5321       if toloc and toloc > -1 then
5322         if Babel.locale_props[toloc].lg then
5323           item.lang = Babel.locale_props[toloc].lg
5324           node.set_attribute(item, LOCALE, toloc)
5325         end
5326         if Babel.locale_props[toloc]['/'..item.font] then
5327           item.font = Babel.locale_props[toloc]['/'..item.font]
5328         end
5329         toloc_save = toloc

```

```

5330     end
5331     elseif not inmath and item.id == 7 then
5332         item.replace = item.replace and Babel.locale_map(item.replace)
5333         item.pre      = item.pre and Babel.locale_map(item.pre)
5334         item.post     = item.post and Babel.locale_map(item.post)
5335     elseif item.id == node.id'math' then
5336         inmath = (item.subtype == 0)
5337     end
5338 end
5339 return head
5340 end
5341 }

```

The code for `\babelcharproperty` is straightforward. Just note the modified lua table can be different.

```

5342 \newcommand\babelcharproperty[1]{%
5343   \count@=#1\relax
5344   \ifvmode
5345     \expandafter\bbl@chprop
5346   \else
5347     \bbl@error{\string\babelcharproperty\space can be used only in\\%
5348               vertical mode (preamble or between paragraphs)}%
5349     {See the manual for futher info}%
5350   \fi}
5351 \newcommand\bbl@chprop[3][\the\count@]{%
5352   \@tempcnta=#1\relax
5353   \bbl@ifunset{\bbl@chprop@#2}%
5354   {\bbl@error{No property named '#2'. Allowed values are\\%
5355             direction (bc), mirror (bmg), and linebreak (lb)}%
5356    {See the manual for futher info}}%
5357   }%
5358   \loop
5359     \bbl@cs{chprop@#2}{#3}%
5360   \ifnum\count@<\@tempcnta
5361     \advance\count@\@ne
5362   \repeat}
5363 \def\bbl@chprop@direction#1{%
5364   \directlua{
5365     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5366     Babel.characters[\the\count@]['d'] = '#1'
5367   }}
5368 \let\bbl@chprop@bc\bbl@chprop@direction
5369 \def\bbl@chprop@mirror#1{%
5370   \directlua{
5371     Babel.characters[\the\count@] = Babel.characters[\the\count@] or {}
5372     Babel.characters[\the\count@]['m'] = '\number#1'
5373   }}
5374 \let\bbl@chprop@bmg\bbl@chprop@mirror
5375 \def\bbl@chprop@linebreak#1{%
5376   \directlua{
5377     Babel.cjk_characters[\the\count@] = Babel.cjk_characters[\the\count@] or {}
5378     Babel.cjk_characters[\the\count@]['c'] = '#1'
5379   }}
5380 \let\bbl@chprop@lb\bbl@chprop@linebreak
5381 \def\bbl@chprop@locale#1{%
5382   \directlua{
5383     Babel.chr_to_loc = Babel.chr_to_loc or {}
5384     Babel.chr_to_loc[\the\count@] =
5385       \bbl@ifblank{#1}{-1000}{\the\bbl@cs{id@#1}}\space

```



```
5386  }}
```

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.

```
5387 \begingroup % TODO - to a lua file
5388 \catcode`\~ = 12
5389 \catcode`\# = 12
5390 \catcode`\% = 12
5391 \catcode`\& = 14
5392 \directlua{
5393   Babel.linebreaking.replacements = {}
5394   Babel.linebreaking.replacements[0] = {} &% pre
5395   Babel.linebreaking.replacements[1] = {} &% post
5396
5397   &% Discretionaries contain strings as nodes
5398   function Babel.str_to_nodes(fn, matches, base)
5399     local n, head, last
5400     if fn == nil then return nil end
5401     for s in string.utfvalues(fn(matches)) do
5402       if base.id == 7 then
5403         base = base.replace
5404       end
5405       n = node.copy(base)
5406       n.char = s
5407       if not head then
5408         head = n
5409       else
5410         last.next = n
5411       end
5412       last = n
5413     end
5414     return head
5415   end
5416
5417   Babel.fetch_subtext = {}
5418
5419   &% TODO. Merge with [1]?? Maybe not - too many differences.
5420   Babel.fetch_subtext[0] = function(head)
5421     local word_string = ''
5422     local word_nodes = {}
5423     local lang
5424     local item = head
5425     local inmath = false
5426
5427     while item do
5428
5429       &% print('++', item)
```

```

5430
5431     if item.id == 11 then
5432         inmath = (item.subtype == 0)
5433     end
5434     if inmath then
5435         goto next
5436     end
5437
5438     if item.id == 29 then
5439         local locale = node.get_attribute(item, Babel.attr_locale)
5440         && print('++', locale)
5441         if lang == locale or lang == nil then
5442             if (item.char ~= 124) then && ie, not | = space
5443                 lang = lang or locale
5444                 word_string = word_string .. unicode.utf8.char(item.char)
5445                 word_nodes[#word_nodes+1] = item
5446             end
5447         else
5448             break
5449         end
5450
5451         elseif item.id == 12 and item.subtype == 13 then
5452             word_string = word_string .. '|'
5453             word_nodes[#word_nodes+1] = item
5454
5455             && Ignore leading unrecognized nodes, too.
5456             elseif word_string ~= '' then
5457                 word_string = word_string .. Babel.us_char
5458                 word_nodes[#word_nodes+1] = item && Will be ignored
5459             end
5460
5461             ::next::
5462             item = item.next
5463         end
5464
5465         && Here and above we remove some trailing chars but not the
5466         && corresponding nodes. But they aren't accessed.
5467         if word_string:sub(-1) == '|' then
5468             word_string = word_string:sub(1,-2)
5469         end
5470         word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5471         return word_string, word_nodes, item, lang
5472     end
5473
5474     Babel.fetch_subtext[1] = function(head)
5475         local word_string = ''
5476         local word_nodes = {}
5477         local lang
5478         local item = head
5479         local inmath = false
5480
5481         while item do
5482
5483             && print('++', item)
5484
5485             if item.id == 11 then
5486                 inmath = (item.subtype == 0)
5487             end
5488             if inmath then

```

```

5489         goto next
5490     end
5491
5492     if item.id == 29
5493         and (item.char ~= 124) &% ie, not |
5494         and (item.char ~= 61) &% ie, not =
5495         and (item.lang == lang or lang == nil) then
5496             lang = lang or item.lang
5497             word_string = word_string .. unicode.utf8.char(item.char)
5498             word_nodes[#word_nodes+1] = item
5499
5500     elseif item.id == 7 and item.subtype == 2 then
5501         word_string = word_string .. '='
5502         word_nodes[#word_nodes+1] = item
5503
5504     elseif item.id == 7 and item.subtype == 3 then
5505         word_string = word_string .. '|'
5506         word_nodes[#word_nodes+1] = item
5507
5508         &% (1) Go to next word if nothing was found, and (2) implicitly
5509         &% remove leading USs.
5510         elseif word_string == '' then
5511             &% pass
5512
5513         &% This is the responsible for splitting by words.
5514         elseif (item.id == 12 and item.subtype == 13) then
5515             break
5516
5517         else
5518             word_string = word_string .. Babel.us_char
5519             word_nodes[#word_nodes+1] = item &% Will be ignored
5520         end
5521
5522         ::next::
5523         item = item.next
5524     end
5525
5526     word_string = unicode.utf8.gsub(word_string, Babel.us_char .. '+$', '')
5527     return word_string, word_nodes, item, lang
5528 end
5529
5530 function Babel.pre_hyphenate_replace(head)
5531     Babel.hyphenate_replace(head, 0)
5532 end
5533
5534 function Babel.post_hyphenate_replace(head)
5535     Babel.hyphenate_replace(head, 1)
5536 end
5537
5538 Babel.us_char = string.char(31)
5539
5540 function Babel.hyphenate_replace(head, mode)
5541     local u = unicode.utf8
5542     local lbkr = Babel.linebreaking.replacements[mode]
5543
5544     local word_head = head
5545
5546     while true do &% for each subtext block
5547

```

```

5548     local w, wn, nw, lang = Babel.fetch_subtext[mode](word_head)
5549
5550     if Babel.debug then
5551         print()
5552         print('@@@@', w, nw)
5553     end
5554
5555     if nw == nil and w == '' then break end
5556
5557     if not lang then goto next end
5558     if not lbkr[lang] then goto next end
5559
5560     %% For each saved (pre|post)hyphenation. TODO. Reconsider how
5561     %% loops are nested.
5562     for k=1, #lbkr[lang] do
5563         local p = lbkr[lang][k].pattern
5564         local r = lbkr[lang][k].replace
5565
5566         if Babel.debug then
5567             print('====', p, mode)
5568         end
5569
5570         %% This variable is set in some cases below to the first *byte*
5571         %% after the match, either as found by u.match (faster) or the
5572         %% computed position based on sc if w has changed.
5573         local last_match = 0
5574
5575         %% For every match.
5576         while true do
5577             if Babel.debug then
5578                 print('-----')
5579             end
5580             local new    %% used when inserting and removing nodes
5581             local refetch = false
5582
5583             local matches = { u.match(w, p, last_match) }
5584             if #matches < 2 then break end
5585
5586             %% Get and remove empty captures (with ()), which return a
5587             %% number with the position), and keep actual captures
5588             %% (from (...)), if any, in matches.
5589             local first = table.remove(matches, 1)
5590             local last  = table.remove(matches, #matches)
5591             %% Non re-fetched substrings may contain \31, which separates
5592             %% subsubstrings.
5593             if string.find(w:sub(first, last-1), Babel.us_char) then break end
5594
5595             local save_last = last %% with A()BC()D, points to D
5596
5597             %% Fix offsets, from bytes to unicode. Explained above.
5598             first = u.len(w:sub(1, first-1)) + 1
5599             last  = u.len(w:sub(1, last-1)) %% now last points to C
5600
5601             if Babel.debug then
5602                 print(p)
5603                 print('', 'sc', 'first', 'last', 'last_m', 'w')
5604             end
5605
5606             %% This loop traverses the matched substring and takes the

```

```

5607      %% corresponding action stored in the replacement list.
5608      %% sc = the position in substr nodes / string
5609      %% rc = the replacement table index
5610      local sc = first-1
5611      local rc = 0
5612      while rc < last-first+1 do %% for each replacement
5613          if Babel.debug then
5614              print('.....')
5615          end
5616          sc = sc + 1
5617          rc = rc + 1
5618          local crep = r[rc]
5619          local char_node = wn[sc]
5620          local char_base = char_node
5621          local end_replacement = false
5622
5623          if crep and crep.data then
5624              char_base = wn[crep.data+first-1]
5625          end
5626
5627          if Babel.debug then
5628              print('*', sc, first, last, last_match, w)
5629          end
5630
5631          if crep and next(crep) == nil then %% {}
5632              last_match = save_last
5633
5634          elseif crep == nil then %% remove
5635              node.remove(head, char_node)
5636              table.remove(wn, sc)
5637              w = u.sub(w, 1, sc-1) .. u.sub(w, sc+1)
5638              last_match = utf8.offset(w, sc)
5639              sc = sc - 1  %% Nothing has been inserted
5640
5641          elseif mode == 1 and crep and (crep.pre or crep.no or crep.post) then
5642              local d = node.new(7, 0)  %% (disc, discretionary)
5643              d.pre = Babel.str_to_nodes(crep.pre, matches, char_base)
5644              d.post = Babel.str_to_nodes(crep.post, matches, char_base)
5645              d.replace = Babel.str_to_nodes(crep.no, matches, char_base)
5646              d.attr = char_base.attr
5647              if crep.pre == nil then  %% TeXbook p96
5648                  d.penalty = crep.penalty or tex.hyphenpenalty
5649              else
5650                  d.penalty = crep.penalty or tex.exhyphenpenalty
5651              end
5652              head, new = node.insert_before(head, char_node, d)
5653              end_replacement = true
5654
5655          elseif crep and crep.penalty then
5656              local d = node.new(14, 0)  %% (penalty, userpenalty)
5657              d.attr = char_base.attr
5658              d.penalty = crep.penalty
5659              head, new = node.insert_before(head, char_node, d)
5660              end_replacement = true
5661
5662          elseif crep and crep.string then
5663              local str = crep.string(matches)
5664              if str == '' then  %% Gather with nil
5665                  refetch = true

```

```

5666         if sc == 1 then
5667             word_head = char_node.next
5668         end
5669         head, new = node.remove(head, char_node)
5670     elseif char_node.id == 29 and u.len(str) == 1 then
5671         char_node.char = string.utfvalue(str)
5672         w = u.sub(w, 1, sc-1) .. str .. u.sub(w, sc+1)
5673         last_match = utf8.offset(w, sc+1)
5674     else
5675         refetch = true
5676         local n
5677         for s in string.utfvalues(str) do
5678             if char_node.id == 7 then
5679                 %% TODO. Remove this limitation.
5680                 texio.write_nl('Automatic hyphens cannot be replaced, just removed.')
5681             else
5682                 n = node.copy(char_base)
5683             end
5684             n.char = s
5685             if sc == 1 then
5686                 head, new = node.insert_before(head, char_node, n)
5687                 word_head = new
5688             else
5689                 node.insert_before(head, char_node, n)
5690             end
5691         end
5692         node.remove(head, char_node)
5693     end %% string length
5694 end %% if char and char.string (ie replacement cases)
5695
5696 %% Shared by disc and penalty.
5697 if end_replacement then
5698     if sc == 1 then
5699         word_head = new
5700     end
5701     if crep.insert then
5702         last_match = save_last
5703     else
5704         node.remove(head, char_node)
5705         w = u.sub(w, 1, sc-1) .. Babel.us_char .. u.sub(w, sc+1)
5706         last_match = utf8.offset(w, sc)
5707     end
5708 end
5709 end %% for each replacement
5710
5711 if Babel.debug then
5712     print('/', sc, first, last, last_match, w)
5713 end
5714
5715 %% TODO. refetch must be eventually unnecessary.
5716 if refetch then
5717     w, wn, nw, lang = Babel.fetch_subtext[mode](word_head)
5718 end
5719
5720 end %% for match
5721 end %% for patterns
5722
5723 ::next::
5724 word_head = nw

```

```

5725     end &% for substring
5726     return head
5727 end
5728
5729 &% This table stores capture maps, numbered consecutively
5730 Babel.capture_maps = {}
5731
5732 &% The following functions belong to the next macro
5733 function Babel.capture_func(key, cap)
5734     local ret = "[" .. cap:gsub('{{[0-9]}}', "")..m[1]..["] .. "]"
5735     ret = ret:gsub('{{[0-9]}|([^\]|+)|(.-)}', Babel.capture_func_map)
5736     ret = ret:gsub("%[%[%]]%.", '')
5737     ret = ret:gsub("%.%.%[%[%]]%", '')
5738     return key .. "[=function(m) return ]] .. ret .. [[ end]]
5739 end
5740
5741 function Babel.capt_map(from, mapno)
5742     return Babel.capture_maps[mapno][from] or from
5743 end
5744
5745 &% Handle the {n|abc|ABC} syntax in captures
5746 function Babel.capture_func_map(capno, from, to)
5747     local froms = {}
5748     for s in string.utfcharacters(from) do
5749         table.insert(froms, s)
5750     end
5751     local cnt = 1
5752     table.insert(Babel.capture_maps, {})
5753     local mlen = table.getn(Babel.capture_maps)
5754     for s in string.utfcharacters(to) do
5755         Babel.capture_maps[mlen][froms[cnt]] = s
5756         cnt = cnt + 1
5757     end
5758     return "]]..Babel.capt_map(m[" .. capno .. "], " ..
5759         (mlen) .. ").. " .. "["
5760 end
5761 }

```

Now the T_EX high level interface, which requires the function defined above for converting strings to functions returning a string. These functions handle the `{n}` syntax. For example, `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).

```

5762 \catcode`\#=6
5763 \gdef\babelposthyphenation#1#2#3{&%
5764     \bbl@activateposthyphen
5765     \begingroup
5766         \def\babeltempa{\bbl@add@list\babeltempb}&%
5767         \let\babeltempb\@empty
5768         \bbl@foreach{#3}{&%
5769             \bbl@ifsamestring{##1}{remove}&%
5770             {\bbl@add@list\babeltempb{nil}}&%
5771             {\directlua{

```

```

5772         local rep = {[#1]}
5773         rep = rep:gsub('^%s*(insert)%s*', 'insert = true, ')
5774         rep = rep:gsub(' (no)%s*=%s*([^\s,]*)', Babel.capture_func)
5775         rep = rep:gsub(' (pre)%s*=%s*([^\s,]*)', Babel.capture_func)
5776         rep = rep:gsub(' (post)%s*=%s*([^\s,]*)', Babel.capture_func)
5777         rep = rep:gsub('(string)%s*=%s*([^\s,]*)', Babel.capture_func)
5778         tex.print([[string\babeltempa{}}] .. rep .. [{}]])
5779     }&%
5780 \directlua{
5781     local lbkr = Babel.linebreaking.replacements[1]
5782     local u = unicode.utf8
5783     &% Convert pattern:
5784     local patt = string.gsub([==[#2]==], '%s', '')
5785     if not u.find(patt, '()', nil, true) then
5786         patt = '()' .. patt .. '()'
5787     end
5788     patt = string.gsub(patt, '%(%)%^\^', '^()')
5789     patt = string.gsub(patt, '%$(%)%', '()$')
5790     patt = u.gsub(patt, '{(.)}',
5791         function (n)
5792             return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5793         end)
5794     lbkr[\the\csname l@#1\endcsname] = lbkr[\the\csname l@#1\endcsname] or {}
5795     table.insert(lbkr[\the\csname l@#1\endcsname],
5796         { pattern = patt, replace = { \babeltempb } })
5797 }&%
5798 \endgroup}
5799 % TODO. Copypaste pattern.
5800 \gdef\babelprehyphenation#1#2#3{&%
5801 \bbl@activateprehyphen
5802 \begingroup
5803 \def\babeltempa{\bbl@add@list\babeltempb}&%
5804 \let\babeltempb\@empty
5805 \bbl@foreach{#3}{&%
5806 \bbl@ifsamestring{##1}{remove}&%
5807 {\bbl@add@list\babeltempb{nil}}&%
5808 {\directlua{
5809     local rep = {[#1]}
5810     rep = rep:gsub('^%s*(insert)%s*', 'insert = true, ')
5811     rep = rep:gsub('(string)%s*=%s*([^\s,]*)', Babel.capture_func)
5812     tex.print([[string\babeltempa{}}] .. rep .. [{}]])
5813 }&%
5814 \directlua{
5815     local lbkr = Babel.linebreaking.replacements[0]
5816     local u = unicode.utf8
5817     &% Convert pattern:
5818     local patt = string.gsub([==[#2]==], '%s', '')
5819     if not u.find(patt, '()', nil, true) then
5820         patt = '()' .. patt .. '()'
5821     end
5822     &% patt = string.gsub(patt, '%(%)%^\^', '^()')
5823     &% patt = string.gsub(patt, '([^\s,]*)%$(%)', '%1()$')
5824     patt = u.gsub(patt, '{(.)}',
5825         function (n)
5826             return '%' .. (tonumber(n) and (tonumber(n)+1) or n)
5827         end)
5828     lbkr[\the\csname bbl@id@@#1\endcsname] = lbkr[\the\csname bbl@id@@#1\endcsname] or {}
5829     table.insert(lbkr[\the\csname bbl@id@@#1\endcsname],
5830         { pattern = patt, replace = { \babeltempb } })

```



```

5831 }&%
5832 \endgroup}
5833 \endgroup
5834 \def\bbl@activateposthyphen{%
5835 \let\bbl@activateposthyphen\relax
5836 \directlua{
5837 Babel.linebreaking.add_after(Babel.post_hyphenate_replace)
5838 }}
5839 \def\bbl@activateprehyphen{%
5840 \let\bbl@activateprehyphen\relax
5841 \directlua{
5842 Babel.linebreaking.add_before(Babel.pre_hyphenate_replace)
5843 }}

```

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`, `hhline`, `colortbl`, `longtable`, `booktabs`, etc. However, `dcolumn` still fails.

```

5844 \bbl@trace{Redefinitions for bidi layout}
5845 \ifx\@eqnnum\@undefined\else
5846 \ifx\bbl@attr@dir\@undefined\else
5847 \edef\@eqnnum{%
5848 \unexpanded{\ifcase\bbl@attr@dir\else\bbl@textdir\@ne\fi}%
5849 \unexpanded\expandafter{\@eqnnum}}}%
5850 \fi
5851 \fi
5852 \ifx\bbl@opt@layout\@nnil\endinput\fi % if no layout
5853 \ifnum\bbl@bidimode>\z@
5854 \def\bbl@nextfake#1{% non-local changes, use always inside a group!
5855 \bbl@exp{%
5856 \mathdir\the\bodydir
5857 #1% Once entered in math, set boxes to restore values
5858 \<ifmmode>%
5859 \everyvbox{%
5860 \the\everyvbox
5861 \bodydir\the\bodydir
5862 \mathdir\the\mathdir
5863 \everyhbox{\the\everyhbox}%
5864 \everyvbox{\the\everyvbox}}%
5865 \everyhbox{%
5866 \the\everyhbox
5867 \bodydir\the\bodydir
5868 \mathdir\the\mathdir
5869 \everyhbox{\the\everyhbox}%
5870 \everyvbox{\the\everyvbox}}%
5871 \<fi>}}%
5872 \def\@hangfrom#1{%

```

```

5873 \setbox\@tempboxa\hbox{#{#1}}%
5874 \hangindent\wd\@tempboxa
5875 \ifnum\bb1@getluadir{page}=\bb1@getluadir{par}\else
5876 \shapemode\@ne
5877 \fi
5878 \noindent\box\@tempboxa}
5879 \fi
5880 \IfBabelLayout{tabular}
5881 {\let\bb1@OL@tabular\@tabular
5882 \bb1@sreplace\@tabular{$}\bb1@nextfake$}%
5883 \let\bb1@NL@tabular\@tabular
5884 \AtBeginDocument{%
5885 \ifx\bb1@NL@tabular\@tabular\else
5886 \bb1@sreplace\@tabular{$}\bb1@nextfake$}%
5887 \let\bb1@NL@tabular\@tabular
5888 \fi}}
5889 {}
5890 \IfBabelLayout{lists}
5891 {\let\bb1@OL@list\list
5892 \bb1@sreplace\list{\parshape}\bb1@listparshape}%
5893 \let\bb1@NL@list\list
5894 \def\bb1@listparshape#1#2#3{%
5895 \parshape #1 #2 #3 %
5896 \ifnum\bb1@getluadir{page}=\bb1@getluadir{par}\else
5897 \shapemode\tw@
5898 \fi}}
5899 {}
5900 \IfBabelLayout{graphics}
5901 {\let\bb1@pictresetdir\relax
5902 \def\bb1@pictsetdir{%
5903 \ifcase\bb1@thetextdir
5904 \let\bb1@pictresetdir\relax
5905 \else
5906 \textdir TLT\relax
5907 \def\bb1@pictresetdir{\textdir TRT\relax}%
5908 \fi}%
5909 \let\bb1@OL@picture\@picture
5910 \let\bb1@OL@put\put
5911 \bb1@sreplace\@picture{\hskip-}\bb1@pictsetdir\hskip-}%
5912 \def\put(#1,#2)#3{% Not easy to patch. Better redefine.
5913 \killglue
5914 \raise#2\unitlength
5915 \hb@xt@z@{\kern#1\unitlength\bb1@pictresetdir#3}\hss}}%
5916 \AtBeginDocument
5917 {\ifx\tikz@atbegin@node\undefined\else
5918 \let\bb1@OL@pgfpicture\pgfpicture
5919 \bb1@sreplace\pgfpicture{\pgfpicturetrue}%
5920 {\bb1@pictsetdir\pgfpicturetrue}%
5921 \bb1@add\pgfsys@beginpicture{\bb1@pictsetdir}%
5922 \bb1@add\tikz@atbegin@node{\bb1@pictresetdir}%
5923 \fi}}
5924 {}

```

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.

```

5925 \IfBabelLayout{counters}%
5926 {\let\bb1@OL@textsuperscript\textsuperscript
5927 \bb1@sreplace\textsuperscript{\m@th}\m@th\mathdir\pagedir}%

```

```

5928 \let\bbl@latinarabic=\@arabic
5929 \let\bbl@OL@@arabic\@arabic
5930 \def\@arabic#1{\babelsublr{\bbl@latinarabic#1}}%
5931 \@ifpackagewith{babel}{bidi=default}%
5932 {\let\bbl@asciroman=\@roman
5933 \let\bbl@OL@@roman\@roman
5934 \def\@roman#1{\babelsublr{\ensureascii{\bbl@asciroman#1}}}%
5935 \let\bbl@asciiRoman=\@Roman
5936 \let\bbl@OL@@roman\@Roman
5937 \def\@Roman#1{\babelsublr{\ensureascii{\bbl@asciiRoman#1}}}%
5938 \let\bbl@OL@labelenumii\labelenumii
5939 \def\labelenumii{}\theenumii{}%
5940 \let\bbl@OL@p@enumiii\p@enumiii
5941 \def\p@enumiii{\p@enumii}\theenumii{}}{}{}%
5942 <<Footnote changes>>
5943 \IfBabelLayout{footnotes}%
5944 {\let\bbl@OL@footnote\footnote
5945 \BabelFootnote\footnote\language\language{}{}}%
5946 \BabelFootnote\localfootnote\language\language{}{}}%
5947 \BabelFootnote\mainfootnote{}{}{}%
5948 {}

```

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

```

5949 \IfBabelLayout{extras}%
5950 {\let\bbl@OL@underline\underline
5951 \bbl@sreplace\underline{$\@@underline}\bbl@nextfake$\@@underline}%
5952 \let\bbl@OL@LaTeX2e\LaTeX2e
5953 \DeclareRobustCommand{\LaTeXe}{\mbox{\m@th
5954 \if b\expandafter\@car\@series\@nil\boldmath\fi
5955 \babelsublr{%
5956 \LaTeX\kern.15em2\bbl@nextfake$_{\textstyle\varepsilon}$}}}
5957 {}
5958 </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.

```
5959 (*basic-r)
5960 Babel = Babel or {}
5961
5962 Babel.bidi_enabled = true
5963
5964 require('babel-data-bidi.lua')
5965
5966 local characters = Babel.characters
5967 local ranges = Babel.ranges
5968
5969 local DIR = node.id("dir")
5970
5971 local function dir_mark(head, from, to, outer)
5972   dir = (outer == 'r') and 'TLT' or 'TRT' -- ie, reverse
5973   local d = node.new(DIR)
5974   d.dir = '+' .. dir
5975   node.insert_before(head, from, d)
5976   d = node.new(DIR)
5977   d.dir = '-' .. dir
5978   node.insert_after(head, to, d)
5979 end
5980
5981 function Babel.bidi(head, ispar)
5982   local first_n, last_n          -- first and last char with nums
5983   local last_es                  -- an auxiliary 'last' used with nums
5984   local first_d, last_d          -- first and last char in L/R block
5985   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):

```
5986   local strong = ('TRT' == tex.pardir) and 'r' or 'l'
5987   local strong_lr = (strong == 'l') and 'l' or 'r'
5988   local outer = strong
5989
5990   local new_dir = false
5991   local first_dir = false
5992   local inmath = false
```

```

5993
5994 local last_lr
5995
5996 local type_n = ''
5997
5998 for item in node.traverse(head) do
5999
6000   -- three cases: glyph, dir, otherwise
6001   if item.id == node.id'glyph'
6002     or (item.id == 7 and item.subtype == 2) then
6003
6004     local itemchar
6005     if item.id == 7 and item.subtype == 2 then
6006       itemchar = item.replace.char
6007     else
6008       itemchar = item.char
6009     end
6010     local chardata = characters[itemchar]
6011     dir = chardata and chardata.d or nil
6012     if not dir then
6013       for nn, et in ipairs(ranges) do
6014         if itemchar < et[1] then
6015           break
6016         elseif itemchar <= et[2] then
6017           dir = et[3]
6018           break
6019         end
6020       end
6021     end
6022     dir = dir or 'l'
6023     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).

```

6024   if new_dir then
6025     attr_dir = 0
6026     for at in node.traverse(item.attr) do
6027       if at.number == luatexbase.registernumber'bbl@attr@dir' then
6028         attr_dir = at.value % 3
6029       end
6030     end
6031     if attr_dir == 1 then
6032       strong = 'r'
6033     elseif attr_dir == 2 then
6034       strong = 'al'
6035     else
6036       strong = 'l'
6037     end
6038     strong_lr = (strong == 'l') and 'l' or 'r'
6039     outer = strong_lr
6040     new_dir = false
6041   end
6042
6043   if dir == 'nsm' then dir = strong end -- W1

```

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

```

6044     dir_real = dir          -- We need dir_real to set strong below
6045     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:

```

6046     if strong == 'al' then
6047         if dir == 'en' then dir = 'an' end          -- W2
6048         if dir == 'et' or dir == 'es' then dir = 'on' end -- W6
6049         strong_lr = 'r'                             -- W3
6050     end

```

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

```

6051     elseif item.id == node.id'dir' and not inmath then
6052         new_dir = true
6053         dir = nil
6054     elseif item.id == node.id'math' then
6055         inmath = (item.subtype == 0)
6056     else
6057         dir = nil          -- Not a char
6058     end

```

Numbers in R mode. A sequence of <en>, <et>, <an>, <es> and <cs> is typeset (with some rules) in L mode. We store the starting and ending points, and only when anything different is found (including nil, ie, a non-char), the textdir is set. This means you cannot insert, say, a whatsit, but this is what I would expect (with luacolor you may colorize some digits). Anyway, this behavior could be changed with a switch in the future. Note in the first branch only <an> is relevant if <al>.

```

6059     if dir == 'en' or dir == 'an' or dir == 'et' then
6060         if dir ~= 'et' then
6061             type_n = dir
6062         end
6063         first_n = first_n or item
6064         last_n = last_es or item
6065         last_es = nil
6066     elseif dir == 'es' and last_n then -- W3+W6
6067         last_es = item
6068     elseif dir == 'cs' then          -- it's right - do nothing
6069     elseif first_n then -- & if dir = any but en, et, an, es, cs, inc nil
6070         if strong_lr == 'r' and type_n ~= '' then
6071             dir_mark(head, first_n, last_n, 'r')
6072         elseif strong_lr == 'l' and first_d and type_n == 'an' then
6073             dir_mark(head, first_n, last_n, 'r')
6074             dir_mark(head, first_d, last_d, outer)
6075             first_d, last_d = nil, nil
6076         elseif strong_lr == 'l' and type_n ~= '' then
6077             last_d = last_n
6078         end
6079         type_n = ''
6080         first_n, last_n = nil, nil
6081     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:

```

6082     if dir == 'l' or dir == 'r' then
6083         if dir ~= outer then

```

```

6084     first_d = first_d or item
6085     last_d = item
6086     elseif first_d and dir ~= strong_lr then
6087         dir_mark(head, first_d, last_d, outer)
6088         first_d, last_d = nil, nil
6089     end
6090 end

```

Mirroring. Each chunk of text in a certain language is considered a “closed” sequence. If <r on r> and <l on l>, it’s clearly <r> and <l>, resptly, but with other combinations depends on outer. From all these, we select only those resolving <on> → <r>. At the beginning (when last_lr is nil) of an R text, they are mirrored directly.

TODO - numbers in R mode are processed. It doesn’t hurt, but should not be done.

```

6091     if dir and not last_lr and dir ~= 'l' and outer == 'r' then
6092         item.char = characters[item.char] and
6093             characters[item.char].m or item.char
6094     elseif (dir or new_dir) and last_lr ~= item then
6095         local mir = outer .. strong_lr .. (dir or outer)
6096         if mir == 'rrr' or mir == 'lrr' or mir == 'rrl' or mir == 'rlr' then
6097             for ch in node.traverse(node.next(last_lr)) do
6098                 if ch == item then break end
6099                 if ch.id == node.id'glyph' and characters[ch.char] then
6100                     ch.char = characters[ch.char].m or ch.char
6101                 end
6102             end
6103         end
6104     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).

```

6105     if dir == 'l' or dir == 'r' then
6106         last_lr = item
6107         strong = dir_real          -- Don't search back - best save now
6108         strong_lr = (strong == 'l') and 'l' or 'r'
6109     elseif new_dir then
6110         last_lr = nil
6111     end
6112 end

```

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

```

6113     if last_lr and outer == 'r' then
6114         for ch in node.traverse_id(node.id'glyph', node.next(last_lr)) do
6115             if characters[ch.char] then
6116                 ch.char = characters[ch.char].m or ch.char
6117             end
6118         end
6119     end
6120     if first_n then
6121         dir_mark(head, first_n, last_n, outer)
6122     end
6123     if first_d then
6124         dir_mark(head, first_d, last_d, outer)
6125     end

```

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

```

6126     return node.prev(head) or head
6127 end
6128 </basic-r>

```

And here the Lua code for bidi=basic:

```
6129 (*basic)
6130 Babel = Babel or {}
6131
6132 -- eg, Babel.fontmap[1][<prefontid>]=<dirfontid>
6133
6134 Babel.fontmap = Babel.fontmap or {}
6135 Babel.fontmap[0] = {}      -- l
6136 Babel.fontmap[1] = {}      -- r
6137 Babel.fontmap[2] = {}      -- al/an
6138
6139 Babel.bidi_enabled = true
6140 Babel.mirroring_enabled = true
6141
6142 require('babel-data-bidi.lua')
6143
6144 local characters = Babel.characters
6145 local ranges = Babel.ranges
6146
6147 local DIR = node.id('dir')
6148 local GLYPH = node.id('glyph')
6149
6150 local function insert_implicit(head, state, outer)
6151   local new_state = state
6152   if state.sim and state.eim and state.sim ~= state.eim then
6153     dir = ((outer == 'r') and 'TLT' or 'TRT') -- ie, reverse
6154     local d = node.new(DIR)
6155     d.dir = '+' .. dir
6156     node.insert_before(head, state.sim, d)
6157     local d = node.new(DIR)
6158     d.dir = '-' .. dir
6159     node.insert_after(head, state.eim, d)
6160   end
6161   new_state.sim, new_state.eim = nil, nil
6162   return head, new_state
6163 end
6164
6165 local function insert_numeric(head, state)
6166   local new
6167   local new_state = state
6168   if state.san and state.ean and state.san ~= state.ean then
6169     local d = node.new(DIR)
6170     d.dir = '+TLT'
6171     _, new = node.insert_before(head, state.san, d)
6172     if state.san == state.sim then state.sim = new end
6173     local d = node.new(DIR)
6174     d.dir = '-TLT'
6175     _, new = node.insert_after(head, state.ean, d)
6176     if state.ean == state.eim then state.eim = new end
6177   end
6178   new_state.san, new_state.ean = nil, nil
6179   return head, new_state
6180 end
6181
6182 -- TODO - \hbox with an explicit dir can lead to wrong results
6183 -- <R \hbox dir TLT{<R>}> and <L \hbox dir TRT{<L>}>. A small attempt
6184 -- was s made to improve the situation, but the problem is the 3-dir
6185 -- model in babel/Unicode and the 2-dir model in LuaTeX don't fit
```



```

6186 -- well.
6187
6188 function Babel.bidi(head, ispar, hdir)
6189   local d    -- d is used mainly for computations in a loop
6190   local prev_d = ''
6191   local new_d = false
6192
6193   local nodes = {}
6194   local outer_first = nil
6195   local inmath = false
6196
6197   local glue_d = nil
6198   local glue_i = nil
6199
6200   local has_en = false
6201   local first_et = nil
6202
6203   local ATDIR = luatexbase.registernumber'bbl@attr@dir'
6204
6205   local save_outer
6206   local temp = node.get_attribute(head, ATDIR)
6207   if temp then
6208     temp = temp % 3
6209     save_outer = (temp == 0 and 'l') or
6210                  (temp == 1 and 'r') or
6211                  (temp == 2 and 'al')
6212   elseif ispar then      -- Or error? Shouldn't happen
6213     save_outer = ('TRT' == tex.pardir) and 'r' or 'l'
6214   else                  -- Or error? Shouldn't happen
6215     save_outer = ('TRT' == hdir) and 'r' or 'l'
6216   end
6217   -- when the callback is called, we are just _after_ the box,
6218   -- and the textdir is that of the surrounding text
6219   -- if not ispar and hdir ~= tex.textdir then
6220   --   save_outer = ('TRT' == hdir) and 'r' or 'l'
6221   -- end
6222   local outer = save_outer
6223   local last = outer
6224   -- 'al' is only taken into account in the first, current loop
6225   if save_outer == 'al' then save_outer = 'r' end
6226
6227   local fontmap = Babel.fontmap
6228
6229   for item in node.traverse(head) do
6230
6231     -- In what follows, #node is the last (previous) node, because the
6232     -- current one is not added until we start processing the neutrals.
6233
6234     -- three cases: glyph, dir, otherwise
6235     if item.id == GLYPH
6236       or (item.id == 7 and item.subtype == 2) then
6237
6238       local d_font = nil
6239       local item_r
6240       if item.id == 7 and item.subtype == 2 then
6241         item_r = item.replace    -- automatic discs have just 1 glyph
6242       else
6243         item_r = item
6244       end

```

```

6245     local chardata = characters[item_r.char]
6246     d = chardata and chardata.d or nil
6247     if not d or d == 'nsm' then
6248         for nn, et in ipairs(ranges) do
6249             if item_r.char < et[1] then
6250                 break
6251             elseif item_r.char <= et[2] then
6252                 if not d then d = et[3]
6253                 elseif d == 'nsm' then d_font = et[3]
6254                 end
6255                 break
6256             end
6257         end
6258     end
6259     d = d or 'l'
6260
6261     -- A short 'pause' in bidi for mapfont
6262     d_font = d_font or d
6263     d_font = (d_font == 'l' and 0) or
6264             (d_font == 'nsm' and 0) or
6265             (d_font == 'r' and 1) or
6266             (d_font == 'al' and 2) or
6267             (d_font == 'an' and 2) or nil
6268     if d_font and fontmap and fontmap[d_font][item_r.font] then
6269         item_r.font = fontmap[d_font][item_r.font]
6270     end
6271
6272     if new_d then
6273         table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6274         if inmath then
6275             attr_d = 0
6276         else
6277             attr_d = node.get_attribute(item, ATDIR)
6278             attr_d = attr_d % 3
6279         end
6280         if attr_d == 1 then
6281             outer_first = 'r'
6282             last = 'r'
6283         elseif attr_d == 2 then
6284             outer_first = 'r'
6285             last = 'al'
6286         else
6287             outer_first = 'l'
6288             last = 'l'
6289         end
6290         outer = last
6291         has_en = false
6292         first_et = nil
6293         new_d = false
6294     end
6295
6296     if glue_d then
6297         if (d == 'l' and 'l' or 'r') ~= glue_d then
6298             table.insert(nodes, {glue_i, 'on', nil})
6299         end
6300         glue_d = nil
6301         glue_i = nil
6302     end
6303

```

```

6304     elseif item.id == DIR then
6305         d = nil
6306         new_d = true
6307
6308     elseif item.id == node.id'glue' and item.subtype == 13 then
6309         glue_d = d
6310         glue_i = item
6311         d = nil
6312
6313     elseif item.id == node.id'math' then
6314         inmath = (item.subtype == 0)
6315
6316     else
6317         d = nil
6318     end
6319
6320     -- AL <= EN/ET/ES      -- W2 + W3 + W6
6321     if last == 'al' and d == 'en' then
6322         d = 'an'          -- W3
6323     elseif last == 'al' and (d == 'et' or d == 'es') then
6324         d = 'on'          -- W6
6325     end
6326
6327     -- EN + CS/ES + EN      -- W4
6328     if d == 'en' and #nodes >= 2 then
6329         if (nodes[#nodes][2] == 'es' or nodes[#nodes][2] == 'cs')
6330             and nodes[#nodes-1][2] == 'en' then
6331             nodes[#nodes][2] = 'en'
6332         end
6333     end
6334
6335     -- AN + CS + AN          -- W4 too, because uax9 mixes both cases
6336     if d == 'an' and #nodes >= 2 then
6337         if (nodes[#nodes][2] == 'cs')
6338             and nodes[#nodes-1][2] == 'an' then
6339             nodes[#nodes][2] = 'an'
6340         end
6341     end
6342
6343     -- ET/EN                -- W5 + W7->1 / W6->on
6344     if d == 'et' then
6345         first_et = first_et or (#nodes + 1)
6346     elseif d == 'en' then
6347         has_en = true
6348         first_et = first_et or (#nodes + 1)
6349     elseif first_et then    -- d may be nil here !
6350         if has_en then
6351             if last == 'l' then
6352                 temp = 'l'    -- W7
6353             else
6354                 temp = 'en'   -- W5
6355             end
6356         else
6357             temp = 'on'       -- W6
6358         end
6359         for e = first_et, #nodes do
6360             if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6361         end
6362         first_et = nil

```

```

6363     has_en = false
6364 end
6365
6366 if d then
6367     if d == 'al' then
6368         d = 'r'
6369         last = 'al'
6370     elseif d == 'l' or d == 'r' then
6371         last = d
6372     end
6373     prev_d = d
6374     table.insert(nodes, {item, d, outer_first})
6375 end
6376
6377 outer_first = nil
6378
6379 end
6380
6381 -- TODO -- repeated here in case EN/ET is the last node. Find a
6382 -- better way of doing things:
6383 if first_et then      -- dir may be nil here !
6384     if has_en then
6385         if last == 'l' then
6386             temp = 'l'    -- W7
6387         else
6388             temp = 'en'   -- W5
6389         end
6390     else
6391         temp = 'on'       -- W6
6392     end
6393     for e = first_et, #nodes do
6394         if nodes[e][1].id == GLYPH then nodes[e][2] = temp end
6395     end
6396 end
6397
6398 -- dummy node, to close things
6399 table.insert(nodes, {nil, (outer == 'l') and 'l' or 'r', nil})
6400
6401 ----- NEUTRAL -----
6402
6403 outer = save_outer
6404 last = outer
6405
6406 local first_on = nil
6407
6408 for q = 1, #nodes do
6409     local item
6410
6411     local outer_first = nodes[q][3]
6412     outer = outer_first or outer
6413     last = outer_first or last
6414
6415     local d = nodes[q][2]
6416     if d == 'an' or d == 'en' then d = 'r' end
6417     if d == 'cs' or d == 'et' or d == 'es' then d = 'on' end --- W6
6418
6419     if d == 'on' then
6420         first_on = first_on or q
6421     elseif first_on then

```

```

6422     if last == d then
6423         temp = d
6424     else
6425         temp = outer
6426     end
6427     for r = first_on, q - 1 do
6428         nodes[r][2] = temp
6429         item = nodes[r][1]    -- MIRRORING
6430         if Babel.mirroring_enabled and item.id == GLYPH
6431             and temp == 'r' and characters[item.char] then
6432             local font_mode = font.fonts[item.font].properties.mode
6433             if font_mode ~= 'harf' and font_mode ~= 'plug' then
6434                 item.char = characters[item.char].m or item.char
6435             end
6436         end
6437     end
6438     first_on = nil
6439 end
6440
6441     if d == 'r' or d == 'l' then last = d end
6442 end
6443
6444 ----- IMPLICIT, REORDER -----
6445
6446 outer = save_outer
6447 last = outer
6448
6449 local state = {}
6450 state.has_r = false
6451
6452 for q = 1, #nodes do
6453     local item = nodes[q][1]
6454
6455     outer = nodes[q][3] or outer
6456
6457     local d = nodes[q][2]
6458
6459     if d == 'nsm' then d = last end          -- W1
6460     if d == 'en' then d = 'an' end
6461     local isdir = (d == 'r' or d == 'l')
6462
6463     if outer == 'l' and d == 'an' then
6464         state.san = state.san or item
6465         state.ean = item
6466     elseif state.san then
6467         head, state = insert_numeric(head, state)
6468     end
6469
6470     if outer == 'l' then
6471         if d == 'an' or d == 'r' then      -- im -> implicit
6472             if d == 'r' then state.has_r = true end
6473             state.sim = state.sim or item
6474             state.eim = item
6475         elseif d == 'l' and state.sim and state.has_r then
6476             head, state = insert_implicit(head, state, outer)
6477         elseif d == 'l' then
6478             state.sim, state.eim, state.has_r = nil, nil, false
6479         end
6480     end

```

```

6481   else
6482     if d == 'an' or d == 'l' then
6483       if nodes[q][3] then -- nil except after an explicit dir
6484         state.sim = item -- so we move sim 'inside' the group
6485       else
6486         state.sim = state.sim or item
6487       end
6488       state.eim = item
6489     elseif d == 'r' and state.sim then
6490       head, state = insert_implicit(head, state, outer)
6491     elseif d == 'r' then
6492       state.sim, state.eim = nil, nil
6493     end
6494   end
6495
6496   if isdir then
6497     last = d -- Don't search back - best save now
6498   elseif d == 'on' and state.san then
6499     state.san = state.san or item
6500     state.ean = item
6501   end
6502
6503 end
6504
6505 return node.prev(head) or head
6506 end
6507 </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.

```

6508 <nil>
6509 \ProvidesLanguage{nil}[<<date>> <<version>> Nil language]
6510 \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.

```

6511 \ifx\l@nil\undefined
6512   \newlanguage\l@nil

```

```

6513 \namedef{bbl@hyphendata@the\l@nil}{\{}}% Remove warning
6514 \let\bbl@elt\relax
6515 \edef\bbl@languages{% Add it to the list of languages
6516 \bbl@languages\bbl@elt{nil}{\the\l@nil}{\{}}
6517 \fi

```

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

```
6518 \providehyphenmins{\CurrentOption}{\m@ne\m@ne}
```

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

```

\captionnil
\datenil
6519 \let\captionnil\@empty
6520 \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.

```

6521 \ldf@finish{nil}
6522 \</nil>

```

16 Support for Plain \TeX (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 \TeX -format. When asked he responded:

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

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

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

```

6523 (*bplain | blplain)
6524 \catcode`\{=1 % left brace is begin-group character
6525 \catcode`\}=2 % right brace is end-group character
6526 \catcode`\#=6 % hash mark is macro parameter character

```

If a file called `hyphen.cfg` can be found, we make sure that *it* will be read instead of the file `hyphen.tex`. We do this by first saving the original meaning of `\input` (and I use a one letter control sequence for that so as not to waste multi-letter control sequence on this in the format).

```

6527 \openin 0 hyphen.cfg
6528 \ifeof0
6529 \else
6530 \let\input

```

Then `\input` is defined to forget about its argument and load `hyphen.cfg` instead. Once that's done the original meaning of `\input` can be restored and the definition of `\a` can be forgotten.

```
6531 \def\input #1 {%
6532   \let\input\a
6533   \a hyphen.cfg
6534   \let\a\undefined
6535 }
6536 \fi
6537 </bplain | bplain>
```

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

```
6538 <bplain>\a plain.tex
6539 <bplain>\a lplain.tex
```

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

```
6540 <bplain>\def\fmtname{babel-plain}
6541 <bplain>\def\fmtname{babel-lplain}
```

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

16.2 Emulating some \LaTeX features

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

```
6542 <<*Emulate LaTeX>> ≡
6543 % == Code for plain ==
6544 \def\@empty{}
6545 \def\loadlocalcfg#1{%
6546   \openin0#1.cfg
6547   \ifeof0
6548     \closein0
6549   \else
6550     \closein0
6551     {\immediate\write16{*****}%
6552      \immediate\write16{* Local config file #1.cfg used}%
6553      \immediate\write16{*}%
6554     }
6555     \input #1.cfg\relax
6556   \fi
6557   \@endofldf}
```

16.3 General tools

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

```
6558 \long\def\@firstofone#1{#1}
6559 \long\def\@firstoftwo#1#2{#1}
6560 \long\def\@secondoftwo#1#2{#2}
6561 \def\@nnil{\@nil}
6562 \def\@gobbletwo#1#2{}
6563 \def\ifstar#1{\@ifnextchar *{\@firstoftwo{#1}}}
6564 \def\@star@or@long#1{%
6565   \@ifstar
6566   {\let\l@ngrel@x\relax#1}%
6567   {\let\l@ngrel@x\long#1}}
```



```

6568 \let\l@ngrel@x\relax
6569 \def\@car#1#2\@nil{#1}
6570 \def\@cdr#1#2\@nil{#2}
6571 \let\@typeset@protect\relax
6572 \let\protected@edef\edef
6573 \long\def\@gobble#1{}
6574 \edef\@backslashchar{\expandafter\@gobble\string\}
6575 \def\strip@prefix#1>{}
6576 \def\g@addto@macro#1#2{%
6577     \toks@\expandafter{#1#2}%
6578     \xdef#1{\the\toks@}}
6579 \def\@namedef#1{\expandafter\def\csname #1\endcsname}
6580 \def\@nameuse#1{\csname #1\endcsname}
6581 \def\@ifundefined#1{%
6582     \expandafter\ifx\csname#1\endcsname\relax
6583     \expandafter\@firstoftwo
6584     \else
6585     \expandafter\@secondoftwo
6586     \fi}
6587 \def\@expandtwoargs#1#2#3{%
6588     \edef\reserved@a{\noexpand#1{#2}{#3}}\reserved@a}
6589 \def\zap@space#1 #2{%
6590     #1%
6591     \ifx#2\@empty\else\expandafter\zap@space\fi
6592     #2}
6593 \let\bbl@trace\@gobble

```

$\text{\LaTeX} 2_{\epsilon}$ has the command `\@onlypreamble` which adds commands to a list of commands that are no longer needed after `\begin{document}`.

```

6594 \ifx\@preamblecmds\@undefined
6595     \def\@preamblecmds{}
6596 \fi
6597 \def\@onlypreamble#1{%
6598     \expandafter\gdef\expandafter\@preamblecmds\expandafter{%
6599         \@preamblecmds\do#1}}
6600 \@onlypreamble\@onlypreamble

```

Mimick \LaTeX 's `\AtBeginDocument`; for this to work the user needs to add `\begin{document}` to his file.

```

6601 \def\begin{document}{%
6602     \@begin{document}hook
6603     \global\let\@begin{document}hook\@undefined
6604     \def\do##1{\global\let##1\@undefined}%
6605     \@preamblecmds
6606     \global\let\do\noexpand}
6607 \ifx\@begin{document}hook\@undefined
6608     \def\@begin{document}hook{}
6609 \fi
6610 \@onlypreamble\@begin{document}hook
6611 \def\AtBeginDocument{\g@addto@macro\@begin{document}hook}

```

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

```

6612 \def\AtEndOfPackage#1{\g@addto@macro\@endoflfd{#1}}
6613 \@onlypreamble\AtEndOfPackage
6614 \def\@endoflfd{}
6615 \@onlypreamble\@endoflfd
6616 \let\bbl@afterlang\@empty
6617 \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.

```
6618 \catcode`\&=\z@
6619 \ifx&\if@files\@undefined
6620 \expandafter\let\csname if@files\expandafter\endcsname
6621 \csname iffalse\endcsname
6622 \fi
6623 \catcode`\&=4
```

Mimick \LaTeX 's commands to define control sequences.

```
6624 \def\newcommand{\@star@or@long\new@command}
6625 \def\new@command#1{%
6626 \@testopt{\@newcommand#1}0}
6627 \def\@newcommand#1[#2]{%
6628 \ifnextchar [{\@xargdef#1[#2]}%
6629 {\@argdef#1[#2]}}
6630 \long\def\@argdef#1[#2]#3{%
6631 \@yargdef#1\@ne{#2}{#3}}
6632 \long\def\@xargdef#1[#2][#3]#4{%
6633 \expandafter\def\expandafter#1\expandafter{%
6634 \expandafter\@protected@testopt\expandafter #1%
6635 \csname\string#1\expandafter\endcsname{#3}}%
6636 \expandafter\@yargdef \csname\string#1\endcsname
6637 \tw@{#2}{#4}}
6638 \long\def\@yargdef#1#2#3{%
6639 \@tempcnta#3\relax
6640 \advance \@tempcnta \@ne
6641 \let\@hash@\relax
6642 \edef\reserved@a{\ifx#2\tw@ [\@hash@1]\fi}%
6643 \@tempcntb #2%
6644 \@whilenum\@tempcntb <\@tempcnta
6645 \do{%
6646 \edef\reserved@a{\reserved@a\@hash@the\@tempcntb}%
6647 \advance\@tempcntb \@ne}%
6648 \let\@hash@###
6649 \l@ngrelx\expandafter\def\expandafter#1\reserved@a}
6650 \def\providecommand{\@star@or@long\provide@command}
6651 \def\provide@command#1{%
6652 \begingroup
6653 \escapechar\m@ne\def\@gtempa{\string#1}%
6654 \endgroup
6655 \expandafter\ifundefined\@gtempa
6656 {\def\reserved@a{\newcommand#1}}%
6657 {\let\reserved@a\relax
6658 \def\reserved@a{\newcommand\reserved@a}}%
6659 \reserved@a}%
6660 \def\DeclareRobustCommand{\@star@or@long\declare@robustcommand}
6661 \def\declare@robustcommand#1{%
6662 \edef\reserved@a{\string#1}%
6663 \def\reserved@b{#1}%
6664 \edef\reserved@b{\expandafter\strip@prefix\meaning\reserved@b}%
6665 \edef#1{%
6666 \ifx\reserved@a\reserved@b
6667 \noexpand\x@protect
6668 \noexpand#1%
6669 \fi
6670 \noexpand\protect
```

```

6671 \expandafter\noexpand\csname
6672 \expandafter\@gobble\string#1 \endcsname
6673 }%
6674 \expandafter\new@command\csname
6675 \expandafter\@gobble\string#1 \endcsname
6676 }
6677 \def\x@protect#1{%
6678 \ifx\protect\@typeset@protect\else
6679 \x@protect#1%
6680 \fi
6681 }
6682 \catcode\&=\z@ % Trick to hide conditionals
6683 \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`.

```

6684 \def\bbl@tempa{\csname newif\endcsname&ifin@}
6685 \catcode\&=4
6686 \ifx\in@\@undefined
6687 \def\in@#1#2{%
6688 \def\in@@##1##2##3\in@@{%
6689 \ifx\in@##2\in@false\else\in@true\fi}%
6690 \in@@#2#1\in@\in@@}
6691 \else
6692 \let\bbl@tempa\@empty
6693 \fi
6694 \bbl@tempa

```

\LaTeX has a macro to check whether a certain package was loaded with specific options. The command has two extra arguments which are code to be executed in either the true or false case. This is used to detect whether the document needs one of the accents to be activated (activegrave and activeacute). For plain \TeX we assume that the user wants them to be active by default. Therefore the only thing we do is execute the third argument (the code for the true case).

```

6695 \def\@ifpackagewith#1#2#3#4{#3}

```

The \LaTeX macro `\@ifl@aded` checks whether a file was loaded. This functionality is not needed for plain \TeX but we need the macro to be defined as a no-op.

```

6696 \def\@ifl@aded#1#2#3#4{}

```

For the following code we need to make sure that the commands `\newcommand` and `\providecommand` exist with some sensible definition. They are not fully equivalent to their \LaTeX 2_ϵ versions; just enough to make things work in plain \TeX environments.

```

6697 \ifx\@tempcnta\@undefined
6698 \csname newcount\endcsname\@tempcnta\relax
6699 \fi
6700 \ifx\@tempcntb\@undefined
6701 \csname newcount\endcsname\@tempcntb\relax
6702 \fi

```

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

```

6703 \ifx\bye\@undefined
6704 \advance\count10 by -2\relax
6705 \fi
6706 \ifx\@ifnextchar\@undefined

```

```

6707 \def\@ifnextchar#1#2#3{%
6708   \let\reserved@d=#1%
6709   \def\reserved@a{#2}\def\reserved@b{#3}%
6710   \futurelet\@let@token\@ifnch}
6711 \def\@ifnch{%
6712   \ifx\@let@token@sptoken
6713     \let\reserved@c\@xifnch
6714   \else
6715     \ifx\@let@token\reserved@d
6716       \let\reserved@c\reserved@a
6717     \else
6718       \let\reserved@c\reserved@b
6719     \fi
6720   \fi
6721   \reserved@c}
6722 \def\:{\let\@sptoken= } \: % this makes \@sptoken a space token
6723 \def\:{\@xifnch} \expandafter\def\:{\futurelet\@let@token\@ifnch}
6724 \fi
6725 \def\@testopt#1#2{%
6726   \@ifnextchar[#{1}{#1[#2]}}
6727 \def\@protected@testopt#1{%
6728   \ifx\protect\@typeset@protect
6729     \expandafter\@testopt
6730   \else
6731     \@x@protect#1%
6732   \fi}
6733 \long\def\@whilenum#1\do #2{\ifnum #1\relax #2\relax\@iwhilenum{#1\relax
6734   #2\relax}\fi}
6735 \long\def\@iwhilenum#1{\ifnum #1\expandafter\@iwhilenum
6736   \else\expandafter\@gobble\fi{#1}}

```

16.4 Encoding related macros

Code from `ltoutenc.dtx`, adapted for use in the plain $\text{T}_{\text{E}}\text{X}$ environment.

```

6737 \def\DeclareTextCommand{%
6738   \@dec@text@cmd\providecommand
6739 }
6740 \def\ProvideTextCommand{%
6741   \@dec@text@cmd\providecommand
6742 }
6743 \def\DeclareTextSymbol#1#2#3{%
6744   \@dec@text@cmd\chardef#1{#2}#3\relax
6745 }
6746 \def\@dec@text@cmd#1#2#3{%
6747   \expandafter\def\expandafter#2%
6748     \expandafter{%
6749       \csname#3-cmd\expandafter\endcsname
6750       \expandafter#2%
6751       \csname#3\string#2\endcsname
6752     }%
6753 %   \let\@ifdefinable\@rc@ifdefinable
6754   \expandafter#1\csname#3\string#2\endcsname
6755 }
6756 \def\@current@cmd#1{%
6757   \ifx\protect\@typeset@protect\else
6758     \noexpand#1\expandafter\@gobble
6759   \fi
6760 }

```

```

6761 \def\@changed@cmd#1#2{%
6762   \ifx\protect\@typeset@protect
6763     \expandafter\ifx\csname\cf@encoding\string#1\endcsname\relax
6764       \expandafter\ifx\csname ?\string#1\endcsname\relax
6765         \expandafter\def\csname ?\string#1\endcsname{%
6766           \@changed@x@err{#1}%
6767         }%
6768       \fi
6769     \global\expandafter\let
6770     \csname\cf@encoding \string#1\expandafter\endcsname
6771     \csname ?\string#1\endcsname
6772   \fi
6773   \csname\cf@encoding\string#1%
6774   \expandafter\endcsname
6775 \else
6776   \noexpand#1%
6777 \fi
6778 }
6779 \def\@changed@x@err#1{%
6780   \errhelp{Your command will be ignored, type <return> to proceed}%
6781   \errmessage{Command \protect#1 undefined in encoding \cf@encoding}}
6782 \def\DeclareTextCommandDefault#1{%
6783   \DeclareTextCommand#1?%
6784 }
6785 \def\ProvideTextCommandDefault#1{%
6786   \ProvideTextCommand#1?%
6787 }
6788 \expandafter\let\csname OT1-cmd\endcsname\@current@cmd
6789 \expandafter\let\csname?-cmd\endcsname\@changed@cmd
6790 \def\DeclareTextAccent#1#2#3{%
6791   \DeclareTextCommand#1{#2}[1]{\accent#3 #1}
6792 }
6793 \def\DeclareTextCompositeCommand#1#2#3#4{%
6794   \expandafter\let\expandafter\reserved@a\csname#2\string#1\endcsname
6795   \edef\reserved@b{\string##1}%
6796   \edef\reserved@c{%
6797     \expandafter\@strip@args\meaning\reserved@a:-\@strip@args}%
6798   \ifx\reserved@b\reserved@c
6799     \expandafter\expandafter\expandafter\ifx
6800     \expandafter\@car\reserved@a\relax\relax\@nil
6801     \@text@composite
6802   \else
6803     \edef\reserved@b##1{%
6804       \def\expandafter\noexpand
6805       \csname#2\string#1\endcsname####1{%
6806         \noexpand\@text@composite
6807         \expandafter\noexpand\csname#2\string#1\endcsname
6808         ####1\noexpand\@empty\noexpand\@text@composite
6809         {##1}%
6810       }%
6811     }%
6812     \expandafter\reserved@b\expandafter{\reserved@a{##1}}%
6813   \fi
6814   \expandafter\def\csname\expandafter\string\csname
6815     #2\endcsname\string#1-\string#3\endcsname{#4}
6816 \else
6817   \errhelp{Your command will be ignored, type <return> to proceed}%
6818   \errmessage{\string\DeclareTextCompositeCommand\space used on
6819     inappropriate command \protect#1}

```

```

6820 \fi
6821 }
6822 \def\@text@composite#1#2#3\@text@composite{%
6823 \expandafter\@text@composite@x
6824 \csname\string#1-\string#2\endcsname
6825 }
6826 \def\@text@composite@x#1#2{%
6827 \ifx#1\relax
6828 #2%
6829 \else
6830 #1%
6831 \fi
6832 }
6833 %
6834 \def\@strip@args#1:#2-#3\@strip@args{#2}
6835 \def\DeclareTextComposite#1#2#3#4{%
6836 \def\reserved@a{\DeclareTextCompositeCommand#1{#2}{#3}}%
6837 \bgroup
6838 \lccode\@=#4%
6839 \lowercase{%
6840 \egroup
6841 \reserved@a @%
6842 }%
6843 }
6844 %
6845 \def\UseTextSymbol#1#2{#2}
6846 \def\UseTextAccent#1#2#3{}
6847 \def\@use@text@encoding#1{}
6848 \def\DeclareTextSymbolDefault#1#2{%
6849 \DeclareTextCommandDefault#1{\UseTextSymbol{#2}#1}%
6850 }
6851 \def\DeclareTextAccentDefault#1#2{%
6852 \DeclareTextCommandDefault#1{\UseTextAccent{#2}#1}%
6853 }
6854 \def\cf@encoding{OT1}

```

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

```

6855 \DeclareTextAccent{"}{OT1}{127}
6856 \DeclareTextAccent{'}{OT1}{19}
6857 \DeclareTextAccent{^}{OT1}{94}
6858 \DeclareTextAccent{\`}{OT1}{18}
6859 \DeclareTextAccent{\~}{OT1}{126}

```

The following control sequences are used in `babel.def` but are not defined for `PLAIN TEX`.

```

6860 \DeclareTextSymbol{\textquotedblleft}{OT1}{92}
6861 \DeclareTextSymbol{\textquotedblright}{OT1}{`\"}
6862 \DeclareTextSymbol{\textquoteleft}{OT1}{`\'}
6863 \DeclareTextSymbol{\textquoteright}{OT1}{`\'}
6864 \DeclareTextSymbol{\i}{OT1}{16}
6865 \DeclareTextSymbol{\ss}{OT1}{25}

```

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

```

6866 \ifx\scriptsize\@undefined
6867 \let\scriptsize\sevenrm
6868 \fi
6869 % End of code for plain

```

6870 <</Emulate LaTeX>

A proxy file:

6871 <plain>

6872 \input babel.def

6873 </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 $T_{\text{E}}\text{X}$ book*, Addison-Wesley, 1986.
- [5] Jukka K. Korpela, *Unicode Explained*, O'Reilly, 2006.
- [6] Leslie Lamport, *\LaTeX , A document preparation System*, Addison-Wesley, 1986.
- [7] Leslie Lamport, in: $T_{\text{E}}\text{X}$ hax Digest, Volume 89, #13, 17 February 1989.
- [8] Ken Lunde, *CJKV Information Processing*, O'Reilly, 2nd ed., 2009.
- [9] Hubert Partl, *German $T_{\text{E}}\text{X}$* , *TUGboat* 9 (1988) #1, p. 70–72.
- [10] Joachim Schrod, *International \LaTeX is ready to use*, *TUGboat* 11 (1990) #1, p. 87–90.
- [11] Apostolos Syropoulos, Antonis Tsolomitis and Nick Sofroniu, *Digital typography using \LaTeX* , Springer, 2002, p. 301–373.
- [12] K.F. Treebus. *Tekstwijzer, een gids voor het grafisch verwerken van tekst*, SDU Uitgeverij ('s-Gravenhage, 1988).