

Π Ο Λ Υ Γ Λ Ω Σ Σ Ι Α

Polyglossia: A Babel Replacement for X_YTeX

FRANÇOIS CHARETTE

2009/01/25 v1.0.2

(PDF file generated on 25th January 2009)

Contents

1	Introduction	2
2	Loading language definition files	3
2.1	The recommended way	3
2.2	The “Babel way”	3
2.3	Supported languages	4
3	Language-switching commands	5
3.1	Other commands	5
4	Font setup	6
5	Language-specific options and commands	7
5.1	arabic	7
5.2	english	7
5.3	esperanto	7
5.4	farsi	7
5.5	german	8
5.6	greek	8
5.7	hebrew	9
5.8	lsorbian and usorbian	9
5.9	magyar	9
5.10	russian	9

5.11	sanskrit	9
5.12	serbian	10
5.13	syriac	10
5.14	thai	10
6	Modifying or extending captions and date formats	10
7	Alphabetic numeration in Greek, Arabic, Hebrew, Syriac and Farsi	10
8	Mappings for Arabic, Farsi and Thai numerals	11
9	Calendars	11
9.1	Hebrew calendar (hebrewcal.sty)	11
9.2	Islamic calendar (hijrical.sty)	11
9.3	Farsi (jalālī) calendar (farsical.sty)	12
10	Acknowledgements	12

1 Introduction

Polyglossia is a package for facilitating multilingual typesetting with \LaTeX . Basically, it can be used as a replacement of `babel` for performing the following tasks automatically:

1. Loading the appropriate hyphenation patterns.
2. Setting the script and language tags of the current font (if possible and available), via the package `fontspec`.
3. Switching to a font assigned by the user to a particular script or language.
4. Adjusting some typographical conventions in function of the current language (such as afterindent, frenchindent, spaces before or after punctuation marks, etc.).
5. Redefining all document strings (like “chapter”, “figure”, “bibliography”).
6. Adapting the formatting of dates (for non-gregorian calendars via external packages bundled with polyglossia: currently the hebrew, islamic and farsi calendars are supported).
7. For languages that have their own numeration system, modifying the formatting of numbers appropriately (this also includes redefining the alphabetic sequence for non-Latin alphabets).
8. Ensuring the proper directionality if the document contains languages that are written from right to left (via the package `bidirectional`, available separately).

Several features of `babel` that do not make sense in the \TeX world (like font encodings, shorthands, etc.) are (obviously) not supported. Generally speaking, `polyglossia` aims to remain as compatible as possible with the fundamental features of `babel` while being cleaner, light-weight, and modern. The package `antomega` has been much beneficial in our attempt to reach this objective.

Requirements: The current version of `polyglossia` makes use of some convenient macros defined in the `etoolbox` package by Philipp Lehmann. Being designed specifically for \LaTeX , it obviously also relies on `fontspec` by Will Robertson. For languages written from right to left, it needs the package `bidir` (by the present author). `Polyglossia` also bundles three packages for calendaric computations (`hebrewcal`, `hijrical`, and `farsical`).

2 Loading language definition files

2.1 The recommended way

You can determine the default language by means of the command:

<code>\setdefaultlanguage</code>	<code>\setdefaultlanguage[options]{lang}</code>	
<code>\setmainlanguage</code>	(or equivalently <code>\setmainlanguage</code>). Secondary languages can be loaded with	
<code>\setotherlanguage</code>	<code>\setotherlanguage[options]{lang}</code>	.

These commands have the advantage of being explicit and of allowing to set language-specific options.¹ It is also possible to load a series of secondary languages at once using

<code>\setotherlanguages</code>	<code>\setotherlanguages{lang1, lang2, lang3, ...}</code>	.
	Language-specific options can be set or changed at any time by means of	
<code>\setkeys</code>	<code>\setkeys{lang}{opt1=value1, opt2=value2, ...}</code>	.

2.2 The “Babel way”

As with `babel`, `polyglossia` also allows to load language definition files as package options. In most cases, option `{lang}` will load the file `gloss-{lang}.ldf`. Note however that the *first* language listed in

`\usepackage[lang1, lang2, ...]polyglossia`

will be the default language for the document, which is the opposite convention of `babel`. Note also that this method may not work in some cases, and should be considered deprecated.

¹More on language-specific options below.

2.3 Supported languages

Table 2.3 lists all languages currently supported. Those in red have specific options and/or command that are explained in section 5 below.

albanian	croatian	galician	lsorbian	serbian
amharic ²	czech	german	magyar	slovak
arabic	danish	greek	norsk	slovenian
bahasai	divehi	hebrew	nynorsk	spanish
bahasam	dutch	hindi	polish	swedish
basque	english	icelandic	portuges	syriac
brazil	esperanto	interlingua	romanian	thai
breton	estonian	irish	russian	turkish
bulgarian	farsi	italian	samin	ukrainian
catalan	finnish	latin	sanskrit	usorbian
coptic	french	latvian	scottish	welsh

Table 1: Languages currently supported in [polyglossia](#)

Some options are convenient shortcuts for loading languages with specific options:

- ▶ american = english with option ‘variant=american’
- ▶ USenglish = english with option ‘variant=american’
- ▶ UKenglish = english with option ‘variant=british’
- ▶ british = english with option ‘variant=british’
- ▶ australian = english with option ‘variant=australian’
- ▶ newzealand = english with option ‘variant=newzealand’
- ▶ ogerman = german with option ‘spelling=old’
- ▶ monogreek = greek with option ‘variant=monotonic’ (or ‘mono’)
- ▶ polygreek = greek with option ‘variant=polytonic’ (or ‘poly’)
- ▶ ancientgreek = greek with option ‘variant=ancient’

Another option (turned off by default) is ‘nocalmarks’, which prevents the redefinition of the internal \LaTeX macros `\markboth` and `\markright`.

There is also the option ‘quiet’ which turns off most info messages and some of the warnings issued by \LaTeX , [fontspec](#) and [polyglossia](#).

²New in version 1.0.1. This should be considered an experimental attempt to port the package [ethiop](#). Feedbacks are welcome.

3 Language-switching commands

Whenever a language definition file `gloss-⟨lang⟩.ldf` is loaded, the command `\text{⟨lang⟩}[⟨options⟩]{...}` becomes available for short insertions of text in that language. For example `\textrussian{\today}` yields 25 января 2009 г. Longer passages are better put between the environment `⟨lang⟩` (again with the possibility of setting language options locally. For instance the following allows us to quote the beginning of Homer's *Iliad*:

```
\begin{greek}[variant=ancient]
μήνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε'
ἔθηκε, πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν ἡρώων, αὐτοὺς δὲ ἐλώρια
τεῦχε κύνεσσιν οἰωνοῖσι τε πᾶσι, Διὸς δ' ἐτελείετο βουλή, ἐξ οὗ δὴ τὰ
πρῶτα διαστήτην ἐρίσαντε Ἀτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.
\end{greek}
```

μήνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος οὐλομένην, ἣ μυρὶ Ἀχαιοῖς ἄλγε' ἔθηκε, πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν ἡρώων, αὐτοὺς δὲ ἐλώρια τεῦχε κύνεσσιν οἰωνοῖσι τε πᾶσι, Διὸς δ' ἐτελείετο βουλή, ἐξ οὗ δὴ τὰ πρῶτα διαστήτην ἐρίσαντε Ἀτρεΐδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.

Note that for Arabic one cannot use the environment `arabic`, as `\arabic` is defined internally by \TeX . In this case we need to use the environment `Arabic` instead. This is the beginning of Ibn Khaldūn's *Muqaddima*:

```
\begin{Arabic}
هو إذ الغاية؛ شريف الفوائد، جمّ المذهب، عزيز فنّ التاريخ فنّ أن اعلم
و الملوك سيرهم، في و الأنبياء أخلاقهم، في الأمم من الماضين أحوال على يوقفنا
أحوال في يرومه لمن ذلك في الإقتداء فائدة تتمّ حتّى وسياستهم؛ دولهم في
و الدنيا. الدين
\end{Arabic}
```

اعلم أنّ فنّ التاريخ فنّ عزيز المذهب، جمّ الفوائد، شريف الغاية؛ إذ هو يوقفنا على أحوال الماضين من الأمم في أخلاقهم، والأنبياء في سيرهم، والملوك في دولهم وسياستهم؛ حتّى تتمّ فائدة الإقتداء في ذلك لمن يرومه في أحوال الدين والدنيا.

3.1 Other commands

The following commands are probably of lesser interest to the end user, but ought to be mentioned here.

<code>\selectbackgroundlanguage</code>	▶ <code>\selectbackgroundlanguage</code> : this selects the global font setup and the numeration definitions for the default language.
<code>\resetdefaultlanguage</code>	▶ <code>\resetdefaultlanguage</code> (experimental): completely switches the default language to another one in the middle of a document: <i>this may have adverse effects!</i>
<code>\normalfontlatin</code>	▶ <code>\normalfontlatin</code> : in an environment where <code>\normalfont</code> has been redefined to a non-latin script, this will call the font defined with <code>\setromanfont</code> etc. Likewise it is possible to use <code>\rmfamilylatin</code> , <code>\sffamilylatin</code> , and <code>\ttfamilylatin</code> .
<code>\rmfamilylatin</code>	▶ Some macros defined in <code>babel</code> 's <code>hyphen.cfg</code> (and thus usually compiled into the xelatex format) are redefined, but keep a similar behaviour, namely <code>\selectlanguage</code> , <code>\foreignlanguage</code> , and the environment <code>otherlanguage</code> .
<code>\sffamilylatin</code>	
<code>\ttfamilylatin</code>	
<code>\selectlanguage</code>	Since the Xe _{La} TeX format incorporates <code>babel</code> 's <code>hyphen.cfg</code> , the low-level commands for hyphenation and language switching defined there are also accessible. ³
<code>\foreignlanguage</code>	
<code>otherlanguage</code>	

4 Font setup

With polyglossia it is possible to associate a specific font with any script or language that occurs in the document. That font should always be defined as `\(script) font` or `\(language) font`. For instance, if the default roman font defined by `\setromanfont` does not support Greek, then one can define the font used to display Greek with:

```
\newfontfamily\greekfont[options]{font}.
```

See the [fontspec](#) documentation for more information.

Whenever a new language is activated, `polyglossia` will first check whether a font has been defined for that language or – for languages in non-Latin scripts – for the script it uses. If it is not defined, it will use the currently active font and will attempt to turn on the appropriate OpenType tags for the script and language used, in case these are available in the font, by means of `fontspec`'s `\addfontfeature`. If the current font does not appear to support the script of that language, an error message is displayed.

³The file `hyphen.cfg` (available on the [Xe_{La}TeX subversion repository](#)) is meant to eventually replace `babel`'s `hyphen.cfg`. If you want to experiment with it, rename it into `hyphen.cfg`, copy it to `.../tex/xelatex/polyglossia/` and rebuild the xelatex format.

5 Language-specific options and commands

This section gives a list of all languages for which options and end-user commands are defined. The default value of each option is given in *italic*.

5.1 arabic

Options:

- **calendar** = *islamic* (= hijri) or gregorian
- **locale** = *default*,⁴ *mashriq*,⁵ *libya*, *algeria*, *tunisia*, *morocco*, or *mauritania*. This setting influences the spelling of the month names for the Gregorian calendar, as well as the form of the numerals (unless overridden by the following option).
- **numerals** = *mashriq* or *maghrib* (the latter is the default when locale = *algeria*, *tunisia* or *morocco*)

Commands:

- `\abjad` ▸ `\abjad` and `\abjadmaghribi` (see section 7)
- `\abjadmaghribi`

5.2 english

Options:

- **variant** = *american* (= us), *usmax* (same as ‘american’ but with additional hyphenation patterns), *british* (= uk), *australian* or *newzealand*
- **ordinalmonthday** = *true/false* (true by default only when variant = *british*)

5.3 esperanto

Commands:

- `\hodiaun` ▸ `\hodiaun` and `\hodiaun` are special forms of `\today` (see the [babel](#) document-
`\hodiaun` ation)

5.4 farsi

Options:

- **numerals** = *western* or *eastern*
- **locale** (not yet implemented)
- **calendar** (not yet implemented)

⁴For Egypt, Sudan, Yemen and the Gulf states.

⁵For Iraq, Syria, Jordan, Lebanon and Palestine.

Commands:

`\abjad` ▶ `\abjad` (see section 7)

5.5 `german`

Options:

- ▶ **spelling** = *new* (= 1996) or *old* (= 1901): indicates whether hyphenation patterns for traditional (1901) or reformed (1996) orthography should be used. The latter is the default.
- ▶ **latesthyphen** = *false* or *true*: if this option is set to *true*, the latest (experimental) hyphenation patterns ‘(n)german-x-latest’ will be loaded instead of ‘german’ or ‘ngerman’. NB: This is based on the file `language.dat` that comes with T_EXLive 2008.
- ▶ **babelshorthands** = *false* or *true*: ← if this is turned on, all shorthands defined in the Babel language for fine-tuning the hyphenation of German words are activated. NB: This feature has not been much tested and should be considered experimental.
 - ▶ `"ck` for `ck` to be hyphenated as `k-k`
 - ▶ `"ff` for `ff` to be hyphenated as `ff-f`; this is also available for the letters `l`, `m`, `n`, `p`, `r` and `t`
 - ▶ `"|` disables the ligature at this position
 - ▶ `"-` for an explicit hyphen sign, allowing hyphenation in the rest of the word
 - ▶ `" "` is like `"-`, but produces no hyphen sign (for compound words with an hyphen, e.g., `foo-"bar`)
 - ▶ `"~` for a compound word mark without a breakpoint
 - ▶ `"=` for a compound word mark with a breakpoint, allowing hyphenation in the composing words.

→ 1.0.2: New!

There are also four shorthands for quotation signs:

- ▶ `"`` for German left double quotes („)
- ▶ `"'` for German right double quotes (")
- ▶ `"<` for French left double quotes («)
- ▶ `">` for French right double quotes (»).

5.6 `greek`

Options:

- ▶ **variant** = *monotonic* (= mono), *polytonic* (= poly), or *ancient*
- ▶ **numerals** = *greek* or *arabic*

- **attic** = *false/true*

Commands:

<code>\Greeknnumber</code>	▸ <code>\Greeknnumber</code> and <code>\greeknumber</code> (see section 7).
<code>\greeknumber</code>	▸ The command <code>\atticnumeral</code> (= <code>\atticnum</code>) (activated with the option <code>attic=true</code>), displays numbers using the acrophonic numbering system (defined in the Unicode range U+10140–U+10174). ⁶
<code>\atticnumeral</code>	
<code>\atticnum</code>	

5.7 hebrew

Options:

- **numerals** = *hebrew* or *arabic*
- **calendar** = *hebrew* or *gregorian*

Commands:

<code>\hebrewnumeral</code>	▸ <code>\hebrewnumeral</code> (= <code>\hebrewalph</code>) (see section 7).
<code>\hebrewalph</code>	

5.8 Isorbian and usorbian

Commands:

<code>\oldtoday</code>	▸ <code>\oldtoday</code> : see the babel documentation.
------------------------	--

5.9 magyar

Commands:

<code>\ontoday</code>	▸ <code>\ontoday</code> (= <code>\ondatemagyar</code>): special forms of <code>\today</code> (see the babel documentation).
<code>\ondatemagyar</code>	

5.10 russian

Options:

- **spelling** = *modern* or *old* (for captions and date only, not for hyphenation)

5.11 sanskrit

Options:

- 1.0.2: New!
- **Script** (default = *Devanagari*). ← The value is passed to **fontspec** in cases where `\sanskritfont` or `\devanagarifont` are not defined. This can be useful if you typeset Sanskrit texts in scripts other than Devanagari.

⁶See the documentation of the **xgreek** package for more details.

5.12 serbian

Options:

- **script** = *cyrillic* or *latin*

5.13 syriac

Options:

- **numerals** = *western* (i.e., 1234567890), *eastern* (for which the Oriental Arabic numerals are used: ١٢٣٤٥٦٧٨٩٠), or *abjad* ←.

→ 1.0.1: New!

Commands:

`\abjadsyriac` ▸ `\abjadsyriac` (see section 7)

5.14 thai

Options:

- **numerals** = *thai* or *arabic*

To insert the word breaks, you need to use an external processor. See the documentation to **thai-latex** and the file `testthai.tex` that comes with this package.

6 Modifying or extending captions and date formats

To redefine internal macros, you can use the command `\gappto` from the package **etoolbox**. For compatibility with Babel the command `\addto` is also available with the same effect. For instance, to change the `\chaptername` for language `lingua`, you can do this:

```
\gappto\linguacaptions{\renewcommand{\chaptername}{Caput}}
```

7 Alphabetic numeration in Greek, Arabic, Hebrew, Syriac and Farsi

In certain languages, numbers can be represented by a special alphanumerical notation.⁷ Note that the Hebrew implementation in **polyglossia** is currently less sophisticated than the one in **babel**, where various special cases are taken into account.

`\greeknumeral` The Greek numerals are obtained with `\greeknumeral` (or `\Greeknumeral` in uppercase). Example: `\greeknumeral{1863}` yields $\alpha\omega\xi\gamma'$.

⁷See, e.g., http://en.wikipedia.org/wiki/Greek_numerals, http://en.wikipedia.org/wiki/Abjad_numerals, and http://en.wikipedia.org/wiki/Hebrew_numerals.

`\abjad` The Arabic *abjad* numbers can be generated with the command `\abjad`. Example: `\abjad{1863}` yields غُضْصِج. In the Maghrib the conventions are somewhat different, and the maghribi forms of the *abjad* numerals are obtained with the `\abjadmaghribi` command. Example: `\abjadmaghribi{1863}` yields شُظْظِصِج.

`\hebrewnumeral` Hebrew numerals are generated with the command `\hebrewnumeral`. Example: `\hebrewnumeral{1863}` yields יטוה'א.

Support is also provided for Syriac abjad numerals, which can be generated with `\abjadsyriac`. Example: `\abjadsyriac{463}` yields ܠܡܝ.

8 Mappings for Arabic, Farsi and Thai numerals

Three fontmappings are provided with polyglossia to allow the automatic conversion of the standard Western numerals (0123456789) in their Arabic, Farsi, or Thai forms. To activate them, one should use the fontspec option `Mapping=arabiddigits` (or `farsiddigits` or `thaidigits`). For instance if `\syriacfont` is defined with the option `Mapping=arabiddigits`, then by typing `\textsyriac{2008}` one obtains ٢٠٠٨.

9 Calendars

9.1 Hebrew calendar (hebrewcal.sty)

The package `hebrewcal.sty` is almost a verbatim copy of `hebcals.sty` that comes with `babel`. The command `\Hebrewtoday` formats the current date in the Hebrew calendar (depending of the current writing direction this will automatically set either in Hebrew script or in roman transliteration).

9.2 Islamic calendar (hijrical.sty)

This new package computes dates in the Islamic (Hijra) calendar, which is lunar.⁸ It provides two macros for the end-user. The command

`\HijriFromGregorian` `\HijriFromGregorian{<year>}{<month>}{<day>}`
`\Hijritoday` sets the counters `Hijriday`, `Hijrimonth` and `Hijriyear`. `\Hijritoday` formats the Hijri date for the current day (depending of the current writing direction this is set either in Arabic or in roman transliteration). It also accepts an optional argument to add or subtract a correction (in days) to the date computed by the arithmetical

⁸It makes use of the arithmetical algorithm in chapter 6 of Reingold & Gershowitz, *Calendrical calculation: the Millenium edition* (Cambridge University Press, 2001).

algorithm.⁹ For instance if `\Hijritoday` yields the date “7 Rajab 1429” (which is the date that was displayed on the front page of aljazeera.net on 11th July 2008), `\Hijritoday[1]` would rather print “8 Rajab 1429” (the date indicated the same day on the site gulfnews.com).

9.3 Farsi (jalālī) calendar (farsical.sty)

Again this is taken almost verbatim from `Arabiftoday.sty` (in the [Arabi](#) package), itself a slight modification of the file `ftoday.sty` in Farsi \TeX .¹⁰ Here we have re-named the command `\ftoday` to `\Jalalitoday`. Example: today is 6 Bahman 1387.

10 Acknowledgements

[Polyglossia](#) is notable for being a recycle box of previous contributions by other people. I take this opportunity to thank the following individuals, whose splendid work has made my task almost trivial in comparison: Johannes Braams and the numerous contributors to the [babel](#) package (in particular Boris Lavva and others for its Hebrew support), Alexej Kryukov ([antomega](#)), Will Robertson ([fontspec](#)), Apostolos Syropoulos ([xgreek](#)), Youssef Jabri ([arabi](#)), and Vafa Khalighi ([xepersian](#)). I should also thank other individuals for their assistance in supporting specific languages: Yves Codet (Sanskrit), Zdenek Wagner (Hindi), Mikhal Oren (Hebrew), Sergey Astanin (Russian), and several other members of the $\text{X}_{\text{Y}}\text{T}_{\text{E}}\text{X}$ user community. And of course my gratitude also goes to Jonathan Kew, the formidable author of $\text{X}_{\text{Y}}\text{T}_{\text{E}}\text{X}$!

⁹The Islamic calendar is indeed a purely lunar calendar based on the observation of the first visibility of the lunar crescent at the beginning of the lunar month, so there can be differences between different localities, as well as between civil and religious authorities.

¹⁰I intend to rewrite [farsical](#) from scratch using the algorithm in Reingold & Gershowitz (ref. n. 8).