

Π Ο Λ Υ Γ Λ Ω Σ Σ Ι Α

Polyglossia: A Babel Replacement for X_YLaTeX

FRANÇOIS CHARETTE

2008/07/11 v1.0

(PDF file generated on 16th July 2008)

Contents

1	Introduction	1
2	Loading language definition files	2
2.1	The recommended way	2
2.2	The “Babel way”	2
2.3	Supported languages	3
3	Language-switching commands	3
3.1	Other commands	4
4	Language-specific options and commands	5
4.1	arabic	5
4.2	english	5
4.3	esperanto	6
4.4	farsi	6
4.5	german	6
4.6	greek	6
4.7	hebrew	6
4.8	lsorbian and usorbian	7
4.9	magyar	7
4.10	russian	7
4.11	serbian	7
4.12	syriac	7

4.13	thai	7
5	Alphabetic numeration in Greek, Arabic, Hebrew, Syriac and Farsi	8
6	Calendars	8
6.1	Hebrew calendar (hebrewcal.sty)	8
6.2	Islamic calendar (hijrical.sty)	8
6.3	Farsi (jalālī) calendar (farsical.sty)	9
7	Acknowledgements	9

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 `bidi`, available separately).

Several features of `babel` that do not make sense in the \LaTeX 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 [bidi](#) (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.

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 [4](#) below.

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

¹More on language-specific options below.

albanian	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
croatian	galician	lsorbian	serbian	

Table 1: Languages currently supported in `polyglossia`

- ▶ 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 ‘nolocalmarks’, which prevents the redefinition of the internal \LaTeX macros `\markboth` and `\markright`.

3 Language-switching commands

Whenever a language definition file `gloss-(lang).ldf` is loaded, the command `\text{<lang>}` becomes available for short insertions of text in that language. For example `\textrussian{\today}` yields 16 июля 2008 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}`

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

Note that for Arabic one cannot use the environment `arabic`, as `\arabic` is
 defined internally by \LaTeX . 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 أنّ اعلم
 والملوك سيرهم، في والأنبياء أخلاقهم، في الأمم من الماضين أحوال على يوقفنا
 أحوال في يرومه لمن ذلك في الإقتداء فائدة تتمّ حتّى وسياستهم؛ دولهم في
 والدنيا. 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 nu- meration 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 rede- fined to a non-latin script, this will call the font defined with <code>\setroman- font</code> etc. Likewise it is possible to use <code>\rmfamilylatin</code> , <code>\sffamilylatin</code> , and <code>\ttfamilylatin</code> .
<code>\rmfamilylatin</code>	
<code>\sffamilylatin</code>	
<code>\ttfamilylatin</code>	

`\selectlanguage` `\selectlanguage`, `\foreignlanguage`, and the environment `other language`.
`\foreignlanguage` Since the X_YTeX format incorporates `babel`'s `hyphen. cfg`, the low-level commands
`other language` for hyphenation and language switching defined there are also accessible.²

4 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*.

4.1 arabic

Options:

- **calendar** = *islamic* (= hijri) or *gregorian*
- **locale** = *default*,³ *mashriq*,⁴ *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 5)
`\abjadmaghribi`

4.2 english

Options:

- **variant** = *american* (= us), *british* (= uk), *australian* or *newzealand*
- **ordinalmonthday** = *true/false* (true by default only when variant = *british*)

4.3 esperanto

Commands:

`\hodiaun` ▸ `\hodiaun` and `\hodiaun` are special forms of `\today` (see the `babel` documen-
`\hodiaun`

²The file `hyphen. cfg` (available on the X_YTeX [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.

³For Egypt, Sudan, Yemen and the Gulf states.

⁴For Iraq, Syria, Jordan, Lebanon and Palestine.

tation)

4.4 farsi

Options:

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

Commands:

`\abjad` ▸ `\abjad` (see section 5)

4.5 german

Options:

- **spelling** = *new* or *old*

4.6 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 5).
<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). ⁵
<code>\atticnumeral</code>	
<code>\atticnum</code>	

4.7 hebrew

Options:

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

Commands:

<code>\hebrewnumeral</code>	▸ <code>\hebrewnumeral</code> (= <code>\hebrewalph</code>) (see section 5).
<code>\hebrewalph</code>	

⁵See the documentation of the [xgreek](#) package for more details.

4.8 lsorbian and usorbian

Commands:

`\oldtoday` ▶ `\oldtoday`: see the [babel](#) documentation.

4.9 magyar

Commands:

`\ontoday` ▶ `\ontoday` (= `\onatemagyar`): special forms of `\today` (see the [babel](#) docu-
`\onatemagyar` mentation).

4.10 russian

Options:

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

4.11 serbian

Options:

▶ **script** = *cyrillic* or *latin*

4.12 syriac

Options:

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

Commands:

`\abjadsyriac` ▶ `\abjadsyriac` (see section 5)

4.13 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.

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

In certain languages, numbers can be represented by a special alphanumerical notation.⁶ Note that the Hebrew implementation in `polyglossia` is currently less sophisticated than the one in `babel`, where various special cases are taken into account.

<code>\greeknumeral</code>	The Greek numerals are obtained with <code>\greeknumeral</code> (or <code>\Greeknumeral</code> in uppercase). Example: <code>\greeknumeral{1863}</code> yields $\alpha\omega\xi\gamma'$.
<code>\abjad</code>	The Arabic <i>abjad</i> numbers can be generated with the command <code>\abjad</code> . Example: <code>\abjad{1863}</code> yields غضسج. In the Maghrib the conventions are somewhat different, and the maghribi forms of the <i>abjad</i> numerals are obtained with the <code>\abjadmaghribi</code> command. Example: <code>\abjadmaghribi{1863}</code> yields شظصج.
<code>\abjadmaghribi</code>	
<code>\hebrewnumeral</code>	Hebrew numerals are generated with the command <code>\hebrewnumeral</code> . Example: <code>\hebrewnumeral{1863}</code> yields יטוה'א.
<code>\abjadsyriac</code>	Support is also provided for Syriac abjad numerals, which can be generated with <code>\abjadsyriac</code> . Example: <code>\abjadsyriac{463}</code> yields ܐܘܠܬ.

6 Calendars

6.1 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).

6.2 Islamic calendar (`hijrical.sty`)

This new package computes dates in the Islamic (Hijra) calendar, which is lunar.⁷ 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

⁶See, e.g., http://en.wikipedia.org/wiki/Greek_numerals, http://en.wikipedia.org/wiki/Abjad_numerals, and http://en.wikipedia.org/wiki/Hebrew_numerals.

⁷It makes use of the arithmetical algorithm in chapter 6 of Reingold & Gershowitz, *Calendrical calculation: the Millenium edition* (Cambridge University Press, 2001).

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 algorithm.⁸ 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).

6.3 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 `FarsiTEX`.⁹ Here we have re-named the command `\ftoday` to `\Jalalitoday`. Example: today is 26 Tīr 1387.

7 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), and other members of the X_YT_EX user community. And of course my gratitude also goes to Jonathan Kew, the formidable author of X_YT_EX!

⁸The 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.

⁹I intend to rewrite `farsical` from scratch using the algorithm in Reingold & Gershowitz (ref. n. 7).