Π Ο Λ Υ Γ Λ Ω Σ Σ Ι Α

Polyglossia: A Babel Replacement for XeLaTeX

François Charette

2008/07/26 v1.0.1
(PDF file generated on 27th July 2008)

Contents

1 Introduction

Polyglossia is a package for facilitating multilingual typesetting with XeLaTeX. 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 XeTeX 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 XeLaTeX, 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:

\setdefaultlanguage `\setdefaultlanguage[⟨options⟩]{lang}`

\setmainlanguage (or equivalently `\setmainlanguage`). Secondary languages can be loaded with

\setotherlanguage `\setotherlanguage[⟨options⟩]{lang}` .

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

\setotherlanguages `\setotherlanguages{lang1,lang2,lang3,…}` .

Language-specific options can be set or changed at any time by means of

\setkeys `\setkeys{⟨lang⟩}{opt1=value1,opt2=value2,…}` .

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.

[1] More 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 'nolocalmarks', 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.

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

3 Language-switching commands

\text⟨lang⟩

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 27 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*:

⟨lang⟩

```
\begin{greek}[variant=ancient]
μῆνιν ἄειδε θεὰ Πηληιάδεω Ἀχιλῆος οὐλομένην, ἣ μυρί' Ἀχαιοῖς ἄλγε'
ἔθηκε, πολλὰς δ' ἰφθίμους ψυχὰς Ἄιδι προΐαψεν ἡρώων, αὐτοὺς δὲ ἑλώρια
τεῦχε κύνεσσιν οἰωνοῖσί τε πᾶσι, Διὸς δ' ἐτελείετο βουλή, ἐξ οὗ δὴ τὰ
πρῶτα διαστήτην ἐρίσαντε Ἀτρείδης τε ἄναξ ἀνδρῶν καὶ δῖος Ἀχιλλεύς.
\end{greek}
```

μῆνιν ἄειδε θεὰ Πηληϊάδεω Ἀχιλῆος οὐλομένην, ἣ μυρί' Ἀχαιοῖς ἄλγε' ἔθηκε, πολλὰς δ' ἰφθίμους ψυχὰς Ἄϊδι προΐαψεν ἡρώων, αὐτοὺς δὲ ἑλώρια τεῦχε κύνεσσιν οἰωνοῖσί 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 LaTeX. In this case we need to use the environment `Arabic` instead. This is the beginning of Ibn Khaldūn's *Muqaddima*:

Arabic

```
\begin{Arabic}
هو إذ الغاية؛ شريف الفوائد، جمّ المذهب، عزيز فنّ التاريخ فنّ أنّ اعلم
و الملوك سيرهم، في و الأنبياء أخلاقهم، في الأمم من الماضين أحو ال على يوقفنا
أحو ال 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.

\selectbackgroundlanguage
\resetdefaultlanguage
\normalfontlatin
\rmfamilylatin
\sffamilylatin
\ttfamilylatin
\selectlanguage
\foreignlanguage
otherlanguage

- `\selectbackgroundlanguage`: this selects the global font setup and the numeration definitions for the default language.
- `\resetdefaultlanguage` (experimental): completely switches the default language to another one in the middle of a document: *this may have adverse effects*!
- `\normalfontlatin`: in an environment where `\normalfont` has been redefined to a non-latin script, this will call the font defined with `\setromanfont` etc. Likewise it is possible to use `\rmfamilylatin`, `\sffamilylatin`, and `\ttfamilylatin`.
- Also some macros defined in babel 's `hyphen. cfg` (and thus usually compiled into the xelatex format) are redefined, but keep a similar behaviour, namely `\selectlanguage`, `\foreignlanguage`, and the environment `otherlanguage`.

Since the XeLaTeX format incorporates babel's `hyphen. cfg`, the low-level commands for hyphenation and language switching defined there are also accessible.[3]

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.

[3] The file `hyphen_cfg` (available on the XeTeX 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 overriden by the following option).
- **numerals** = *mashriq* or maghrib (the latter is the default when locale = algeria, tunisia or morocco)

Commands:

`\abjad` `\abjadmaghribi`

- `\abjad` and `\abjadmaghribi` (see section 6)

5.2 english

Options:

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

5.3 esperanto

Commands:

`\hodiau` `\hodiaun`

- `\hodiau` and `\hodiaun` are special forms of `\today` (see the babel documentation)

5.4 farsi

Options:

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

Commands:

`\abjad`

- `\abjad` (see section 6)

[4] For Egypt, Sudan, Yemen and the Golf states.
[5] For Iraq, Syria, Jordan, Lebanon and Palestine.

5.5 german

Options:

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

5.6 greek

Options:

- **variant** = *monotonic* (= mono), polytonic (= poly), or ancient
- **numerals** = *greek* or arabic
- **attic** = *false*/true

Commands:

\Greeknumber
\greeknumber
\atticnumeral
\atticnum

- `\Greeknumber` and `\greeknumber` (see section 6).
- The command `\atticnumeral` (= `\atticnum`) (activated with the option `attic=true`), displays numbers using the acrophonic numbering system (defined in the Unicode range U+10140–U+10174).[6]

5.7 hebrew

Options:

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

Commands:

\hebrewnumeral
\hebrewalph

- `\hebrewnumeral` (= `\hebrewalph`) (see section 6).

5.8 lsorbian and usorbian

Commands:

\oldtoday

- `\oldtoday`: see the babel documentation.

5.9 magyar

Commands:

\ontoday
\ondatemagyar

- `\ontoday` (= `\ondatemagyar`): special forms of `\today` (see the babel documentation).

[6] See the documentation of the xgreek package for more details.

5.10 russian

Options:

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

5.11 serbian

Options:

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

5.12 syriac

Options:

- **numerals** = *western* (i.e., 1234567890), eastern (for which the Oriental Arabic numerals are used: ١٢٣٤٥٦٧٨٩٠), or abjad (new in version 1.0.1).

Commands:

`\abjadsyriac`

- `\abjadsyriac` (see section 6)

5.13 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.

6 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` `\Greeknumeral`

The Greek numerals are obtained with `\greeknumeral` (or `\Greeknumeral` in uppercase). Example: `\greeknumeral{1863}` yields ,αωξγ′.

`\abjad`

The Arabic *abjad* numbers can be generated with the command `\abjad`. Example: `\abjad{1863}` yields غضسجـ. 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 شظصجـ.

`\abjadmaghribi`

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

\hebrewnumeral Hebrew numerals are generated with the command \hebrewnumeral. Example: \hebrewnumeral{1863} yields א׳ח׳סג.

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

7 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=arabicdigits (or farsidigits or thaidigits). For instance if \syriacfont is defined with the option Mapping=arabicdigits, then by typing \textsyriac{2008} one obtains ٢٠٠٨.

8 Calendars

8.1 Hebrew calendar (hebrewcal.sty)

The package hebrewcal.sty is almost a verbatim copy of hebcal.sty that comes
\Hebrewtoday 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).

8.2 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 algorithm.[9] For instance if \Hijritoday yields the date "7 Rajab 1429" (which is

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

[9] The 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 their can be differences between different localities, as well as between civil and religious authorities.

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).

8.3 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.[10] Here we have renamed the command `\ftoday` to `\Jalalitoday`. Example: today is 6 Mordād 1387.

`\Jalalitoday`

9 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 comparision: 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 XƎTEX user community. And of course my gratitude also goes to Jonathan Kew, the formidable author of XƎTEX!

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