

## 2.5.8 The Unicode option

By Axel Kielhorn <[A.Kielhorn@web.de](mailto:A.Kielhorn@web.de)>

Unicode is the way to go if you want to include several languages in one document, especially when these languages are not using the latin script. There are two  $\text{T}_{\text{E}}\text{X}$ -engines that are capable of processing unicode input:

**Xe $\text{T}_{\text{E}}\text{X}$**  was developed for MacOS X but is now available for all architectures. It was first included into TexLive 2007.

**lua $\text{T}_{\text{E}}\text{X}$**  is the successor of pdf $\text{T}_{\text{E}}\text{X}$ . It was first included into TexLive 2009.

The following describes Xe $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  as distributed with TexLive 2010.

### Quickstart

To convert an existing  $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  file to Xe $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  the following needs to be done:

1. Save the file as UTF-8
2. Remove

```
\usepackage{inputenc}
\usepackage{fontenc}
```

from the preamble.

3. Change

```
\usepackage[languageA]{babel}
```

to

```
\usepackage{polyglossia}
\setdefaultlanguage[babelshorthands]{languageA}
```

4. Add

```
\usepackage[Ligatures=TeX]{fontspec}
```

to the preamble.

The package `polyglossia` is a replacement for `babel`. It takes care of the hyphenation patterns and automatically generated text strings. The option `babelshorthands` enables `babel` compatible shorthands for german and catalan.

The package `fontspec` handles font loading for XeTeX and luaTeX. The default font is Latin Modern. It is a little known fact that some TeX commands are ligatures defined in the Computer Modern fonts. The option `Ligatures=TeX` defines the following ligatures:

```
--   -
---  —
’ ’  ”
‘ ‘  “
! ‘  ¡
? ‘  ¿
’ ’  ”
<<  «
>>  »
```

### It’s all $\gamma\rho\eta\eta\kappa$ to me

So far there has been no advantage of using a unicode TeX engine. This changes when we leave the latin script and move to a more interesting language like greek, russian or even hebrew. With a unicode based system, you can simply<sup>12</sup> enter the characters in your editor and TeX will understand them.

Sometimes the font used in the main document does not contain glyphs that are required in the second language<sup>13</sup>. The solution is to define a font that will be used for that language. Whenever a new language is activated, `polyglossia` will first check whether a font has been defined for that language.

```
\newfontfamily\russianfont[Script=Cyrillic,(...)]{(font)}
```

Some languages are written left to right, others are written right to left (RTL). `polyglossia` needs the `bidi` package<sup>14</sup> in order to support RTL languages. The `bidi` package should be the last package you load, even after `hyperref` which is usually the last package.

<sup>12</sup>For small values of simple.

<sup>13</sup>Latin Modern does not contain cyrillic letters

<sup>14</sup>`bidi` does not support luatex.