`hyph-utf8`

Mojca Miklavec (current maintainer)
Arthur Reutenauer (no longer active on hyph-utf8)

April 16, 2012

Abstract

The `hyph-utf8` package gathers all the existing hyphenation patterns for TEX, converted to UTF-8. They can be used directly by UTF-8-aware TEX engines such as LuaTEX and X TEX, and there is a mechanism to convert the patterns to some 8-bit encoding when used with pdfTEX or Knuth's TEX.

List of supported languages

English

-	**english**	usenglish, USenglish, american
en-us	usenglishmax	
en-gb	ukenglish	british, UKenglish

Afrikaans

af	afrikaans

Ancientgreek

grc	ancientgreek
grc-x-ibycus	ibycus

Arabic

ar	arabic

Armenian

hy	armenian

Assamese

as	assamese

Basque

eu	basque

Bengali

bn	bengali

Bulgarian

bg	bulgarian

Catalan

ca	catalan

Chinese

zh-latn-pinyin	pinyin

Coptic

cop	coptic

Croatian

hr	croatian

Czech

cs	czech

Danish

da	danish

Dutch

nl	dutch

Esperanto

eo	esperanto

Estonian

et	estonian

Ethiopic

mul-ethi	ethiopic	amharic, geez

Farsi

fa	farsi	persian

Finnish

fi	finnish

French

fr	french	patois, francais

Friulan

fur	friulan	furlan

Galician

gl	galician

German

de-1901	german
de-1996	ngerman
de-ch-1901	swissgerman

Greek

el-monoton	monogreek	
el-polyton	greek	polygreek

Gujarati

gu	gujarati

Hindi

hi	hindi

Hungarian

hu	hungarian

Icelandic

is	icelandic

Indonesian			
id	indonesian		
Interlingua			
ia	interlingua		
Irish			
ga	irish		
Italian			
it	italian		
Kannada			
kn	kannada		
Kurmanji			
kmr	kurmanji		
Lao			
lo	lao		
Latin			
la	latin		
Latvian			
lv	latvian		
Lithuanian			
lt	lithuanian		
Malayalam			
ml	malayalam		
Marathi			
mr	marathi		
Mongolian			
mn-cyrl	mongolian		
mn-cyrl-x-lmc	mongolianlmc		
Norwegian			
nb	bokmal	norwegian, norsk	
nn	nynorsk		
Oriya			
or	oriya		
Panjabi			
pa	panjabi		
Polish			
pl	polish		
Portuguese			
pt	portuguese	portuges	
Romanian			
ro	romanian		
Romansh			
rm	romansh		
Russian			
ru	russian		
Sanskrit			
sa	sanskrit		
Serbian			
sr-latn	serbian		
sr-cyrl	serbianc		
Slovak			
sk	slovak		
Slovenian			
sl	slovenian	slovene	
Spanish			
es	spanish	espanol	
Swedish			
sv	swedish		
Tamil			
ta	tamil		
Telugu			
te	telugu		
Turkish			
tr	turkish		
Turkmen			
tk	turkmen		
Ukrainian			
uk	ukrainian		
Uppersorbian			
hsb	uppersorbian		
Welsh			
cy	welsh		

Babel defines a few more synonyms (which consequently only work in LaTeX):

english	canadian
british	australian, newzealand
german	austrian
ngerman	naustrian
portuguese	brazilian, brazil

Using hyphenation patterns

Plain TₑX

In engines that support ε-TₑX you can select the desired hyphenation patterns with:

`\uselanguage{langname}`

where `langname` is the string identifying a particular hyphenation file in `language.dat` and can be taken from table on the first two pages.

LᴬTₑX

Since Babel's `hyphen.cfg` is built in the XeLᴬTₑXformat, hyphenation patterns can be used without even loading Babel or Polyglossia. At the low-level this simply corresponds to defining

`\language=\l@<langname>`

The user command is supposed to be

`\hyphenrules{langname}`

or

`\begin{hyphenrules}{langname} ... \end{hyphenrules}.`

and should work with any flavour of LᴬTₑX, however we couldn't make it work.

LᴬTₑX with Babel

You can use Babel with any TₑX engine, however it is currently unmaintained and has never been adapted to work well with Unicode engines. If you are using XₑTₑX please use Polyglossia instead.

`\usepackage[languagename]{babel}`

LᴬTₑX with Polyglossia

Polyglossia should be the preferred choice when using XeLᴬTₑX. It doesn't support LuaLᴬTₑX yet, but it is planned to extend it in future.

```
\usepackage{polyglossia}
\setmainlanguage[optional settings]{langname}
\setotherlanguages{otherlangname}
```

`\begin[optional settings]{otherlangname} ... \end{otherlangname}`

See Polyglossia manual for extensive list of options.

ConTEXt

ConTeXt doesn't load patterns for all the language that hyph-utf8 provides. If you miss any language, please contact the mailing list. The general syntax for supported languages is the following:

```
% language of the main document
\mainlanguage[language]

{\language[otherlanguage] language of some short fragment}
```

You can use full language name or language code. When using ConTEXt MKII you might need to select the appropriate font encoding for Cyrillic scripts, Polish and some other languages:

```
\usetypescript[iwona][qx]
\setupbodyfont[iwona]
\mainlanguage[polish]
```

ConTEXt loads hyphenation patterns in several encodings, so that you can for example use Czech patterns with either `ec` or `il2` font encodings. The right hyphenation patterns will be chosen based on current font encoding.

More examples

Example for Polyglossia

```
\usepackage{polyglossia}
% the language used for main document
\setmainlanguage{asturian}
% American English with extended hyphenation patterns
\setotherlanguage[variant=usmax]{english}
% German with experimental patterns "ngerman-x-latest"
\setotherlanguage[spelling=new,latesthyphen=true]{german}
\setotherlanguages{spanish,catalan,french}

\begin{document}

Long Asturian text ... (Hyphenation for Asturian is not available, but polyglossia automatically
falls back on Catalan for now, which seems to be a reasonable choice.)

\begin{german}
Deutscher Text ... (with the hyphenation patterns selected above: "ngerman-x-latest")
\end{german}

\begin[script=fraktur,spelling=old]{german}
Deutfcher Text ... (set in Fraktur, with traditional hyphenation).
\end{german}

\end{document}
```