hyph-utf8

Mojca Miklavec, Manuel Pégourié-Gonnard, Khaled Hosny (current maintainers) Arthur Reutenauer (no longer active on hyph-utf8)

2013-05-08

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 XeTeX, 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

en-us usenglish usenglish, USenglish, american usenglishmax en-us usenglishmax tritish, UKenglish Afrikaans eo esperanto Ancientgreek gr et estonian gr ancientgreek et etonian gr ancientgreek et etonian gr ancientgreek et ethiopic Arabic ethiopic amharic, gez Arabic Farsi persian Armenian farsi farsi Asmeneian french french as assamese French french Basque fr fruidan bulgarian gl galician Catalan<	English					
en-gb Ukenglish Esperanto Afrikaans eo esperanto af afrikaans eo esperanto Ancientgreek et estonian grc ancientgreek et estonian grc ancientgreek et estonian grc ancientgreek et etonian grc ancientgreek et etonian grc ancientgreek et ethiopic Arabic French French persian Armenian fa farsi persian hy armenian fin finnish Assamese French patois, francais Basque french patois, francais eu basque French Bulgarian Galician French Bulgarian Galician Georgian Catalan Georgian German Chinese German German g	-	english	usenglish, USenglish, ame	erican		
Afrikaans Esperanto af afrikaans e0 esperanto Ancientgreek et estonian grc ancientgreek et estonian grc-x-ibycus ibycus Ethiopic amharic, geez Arabic mul-ethi ethiopic amharic, geez ar arabic Farsi arsia persian Armenian fa fassi persian hy armenian Finnish French French French French Basque French	en-us	usenglishmax				
afikaans eo esperanto Ancientgreek et estonian grc-x-ibycus bycus Ethiopic Arabic mul-ethi ethiopic amharic, geez Arabic Farsi persian Armenian Finnish Assames persian hy armenian Finnish Assames French Basque fr french patois, francais eu basque Friulan patois, francais Bengali fur friulan Friulan Bulgarian Galician Galician Bulgarian Georgian Georgian Catalan German German Chinese de-1901 german zh-latn-pinyin pinyin de-1906 ngerman Copic Greek Greek Copic Croatian el-monoton monogreek polygreek Croatian el-monoton monogreek polygreek Csech <t< td=""><td>en-gb</td><td>ukenglish</td><td>british, UKenglish</td><td></td><td></td><td></td></t<>	en-gb	ukenglish	british, UKenglish			
Ancientgreek et estonian grC ancientgreek et estonian grC×-ibjgus ibjgus Ethiopic Arabic mul-ethi ethiopic amharic, geez ar arabic Farsi persian Armenian fa farsi persian hy armenian Finnish French Sasamese French Sasamese French Basque patois, francais eu basque French French Bengali fur friulan French	Afrikaans			Esperanto		
grc grcx-ribycus ancientgreek ibycus et Ethiopic estonian Arabic mul-ethi ethiopic amharic, geez ar Armenian fa fa fa fa finnish persian Assamese finnish parsian as Basque assamese French patois, francais eu basque frulan patois, francais eu basque Friulan process process Bulgarian Galician gl galician process Catalan German German French	af	afrikaans		eo	esperanto	
Ethiopic Arabic mul-ethi ethiopic amharic, geez ar arabic Farsi Fresh Armenian fa farsi persian hy amenian finnish French Assamese fr french patois, francais eu basque fru friulan patois, francais Bengali fur friulan French French French French French French French French Patois, francais Gethic French French French Patois, francais Gethic French	Ancientgreek			Estonian		
Arabic ar arabic arabic Farsi fa farsi persian Armenian hy armenian Finnish	grc	ancientgreek		et	estonian	
ar arabic Farsi Armenian fa farsi persian hy armenian Finnish Assamese f finnish as assamese French Basque fr french patois, francais eu basque Friulan band fur friulan bn bengali Galician bg bulgarian Georgian Calalan German ca catalan German Chinese de-1901 german zh-latn-pinyin pinyin de-1996 ngerman Coptic Geek Coptic Geek Croatian Greek Croatian el-monoton monogreek hr croatian el-monoton monogreek Croatian el-polyton greek polygreek Czech gu gujarati cs czech gu gujarati da danish hi hindi	grc-x-ibycus	ibycus		Ethiopic		
ArmenianfafasipersianhyarmenianFinnishAssamesefifinnishasassameseFrenchBasquefrfrenchpatois, francaiseubasqueFriulanBengalifurfriulanbnbengaliGalicianBulgarianglgalicianbgbulgarianGeorgiancacatalanGermanChinesede-1901germanzh-latn-pinyinpinyinde-1996ngermanCopticGreekCroatianGeekCroatianel-polytongreekCzechGujaraticsczechgugujaratiDanishHindihindidadanishhihindiHungarianHungarian	Arabic			mul-ethi	ethiopic	amharic, geez
hy armenian Finnish Assamese fi finnish as assamese French Basque fr french patois, francais eu basque Friulan	ar	arabic		Farsi		
Assamese fine finnish as assamese French Basque french eu basque Friulan Bengali fur friulan bn bengali Galician bg bulgarian bg bulgarian Catalan German Chinese zh-latn-pinyin pinyin de-1996 ngerman Coptic Cop coptic Croatian Log coptic Croatian Croatia	Armenian			fa	farsi	persian
as assamese French Basque fr french patois, francais eu basque Friulan Bengali fur friulan bn bengali Galician Bulgarian gl galician bg bulgarian Georgian ca catalan German ca catalan German Zh-latn-pinyin pinyin de-1901 german zh-latn-pinyin pinyin de-ch-1901 swissgerman cop coptic Geek Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech gu gujarati cs czech gu gujarati Danish Hindi hindi da danish hindi hindi	hy	armenian		Finnish		
Basque fr french patois, francais eu basque Friulan Friulan Bengali fur friulan bn bengali Galician Bulgarian gl galician bg bulgarian Georgian ca catalan German Chinese de-1901 german zh-latn-pinyin pinyin de-1996 ngerman cop coptic Greek Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech gu gujarati cs czech gu gujarati Danish Hindi hindi da danish hi midi	Assamese			fi	finnish	
eu basque Friulan Bengali fur friulan bn bengali Galician Bulgarian gl galician bg bulgarian Georgian Catalan German Chinese de-1901 german zh-latn-pinyin pinyin de-1996 ngerman cop coptic Greek Croatian el-monoton monogreek hr croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech gu gujarati Cs czech gu gujarati Danish da danish hi hindi	as	assamese		French		
Bengali fur friulan bn bengali Galician Bulgarian gl galician bg bulgarian Georgian Catalan ka georgian Chinese de-1901 german zh-latn-pinyin pinyin de-1996 ngerman cop coptic Greek Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech Gujarati cs czech gu gujarati Danish da danish hi hindi Dutch	Basque			\mathbf{fr}	french	patois, francais
Bulgarian gl galician bg bulgarian Georgian Catalan ka georgian Chinese de-1901 german zh-latn-pinyin pinyin de-1996 ngerman cop coptic Greek Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech gu gujarati Danish da danish hi hindi Dutch Georgian Georgian Georgian Georgian German German German German German German Gel-1901 german	eu	basque		Friulan		
Bulgarian bg bulgarian Catalan ca catalan Chinese zh-latn-pinyin pinyin cop coptic Croatian cop coptic Croatian croatian croatian cop coptic Croatian cop coptic Croatian cop coptic Croatian bg georgian cop coptic Croatian cop coptic Croatian cop coptic Croatian bg georgian cop coptic cop coptic Croatian cop coptic Croatian cop coptic Croatian cop croatian bg german cop coptic Croatian cop coptic Croatian cop coptic Croatian bli di da danish hi hindi Hungarian	Bengali			fur	friulan	
GeorgianCatalankageorgiancacatalanGermanChinesede-1901germanzh-latn-pinyinpinyinde-1996ngermanCopticde-ch-1901swissgermancopcopticGreekCroatianel-monotonmonogreekhrcroatianel-polytongreekpolygreekCzechGujaraticsczechgugujaratiDanishHindidadanishhihindiDutchHungarian	bn	bengali		Galician		
GeorgianCatalankageorgiancacatalanGermanChinesede-1901germanzh-latn-pinyinpinyinde-1996ngermanCopticde-ch-1901swissgermancopcopticGreekCroatianel-monotonmonogreekhrcroatianel-polytongreekpolygreekCzechgugujaraticsczechgugujaratiDanishHindidadanishhihindiDutchHungarian	Bulgarian	_		gl	galician	
CatalanGermanChinesede-1901germanzh-latn-pinyinpinyinde-1996ngermanCopticde-ch-1901swissgermancopcopticGreekCroatianel-monotonmonogreekhrcroatianel-polytongreekpolygreekCzechGujaraticsczechgugujaratiDanishHindidadanishhihindiDutchHungarian	bg	bulgarian		Georgian		
Chinesede-1901germanzh-latn-pinyinpinyinde-1996ngermanCopticde-ch-1901swissgermancopcopticGreekCroatianel-monotonmonogreekhrcroatianel-polytongreekpolygreekCzechGujaraticsczechgugujaratiDanishHindidadanishhihindiDutchHungarian	Catalan			ka	georgian	
zh-latn-pinyin pinyin de-1996 ngerman Coptic de-ch-1901 swissgerman cop coptic Greek Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech gu gujarati cs czech gu gujarati Danish Hindi da danish hi hindi Dutch Hungarian	ca	catalan		German		
Copticde-ch-1901swissgermancopcopticGreekCroatianel-monotonmonogreekhrcroatianel-polytongreekpolygreekCzechGujaraticsczechgugujaratiDanishHindidadanishhihindiDutchHungarian	Chinese			de-1901	german	
cop coptic Greek Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech Gujarati cs czech gu gujarati Danish Hindi da danish hi hindi Dutch Hungarian	zh-latn-pinyin	pinyin		de-1996	ngerman	
Croatian el-monoton monogreek hr croatian el-polyton greek polygreek Czech Gujarati cs czech gu gujarati Danish Hindi da danish hi hindi Dutch Hungarian	Coptic			de-ch-1901	swissgerman	
hr croatian el-polyton greek polygreek Czech Gujarati cs czech gu gujarati Danish Hindi da danish hi hindi Dutch Hungarian	cop	coptic		Greek		
Czech Gujarati cs czech gu gujarati Danish Hindi da danish hi hindi Dutch Hungarian	Croatian			el-monoton	monogreek	
cs czech gu gujarati Danish Hindi da danish hi hindi Dutch Hungarian	hr	croatian		el-polyton	greek	polygreek
Danish da danish hi hindi Dutch Hungarian	Czech			Gujarati		
da danish hi hindi Dutch Hungarian	CS	czech		gu	gujarati	
Dutch Hungarian	Danish			Hindi		
· · · · · · · · · · · · · · · · · · ·	da	danish		hi	hindi	
nl dutch hu hungarian	Dutch			Hungarian		
	nl	dutch		hu	hungarian	

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

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

english canadian

british australian, newzealand

germanaustrianngermannaustrianportuguesebrazilian, brazil

Using hyphenation patterns

Plain T_FX

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=\l0<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 i12 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}
Deutscher Text ... (set in Fraktur, with traditional hyphenation).
\end{german}
\end{document}
```