Greek Language Support for \LaTeX{} and \LuaLaTeX

Apostolos Syropoulos
Xanthi, Greece
asyropoulos@yahoo.com

2009/11/23
Last Updated 2018/02/16

Abstract
The \texttt{xgreek} package provides rudimentary support for Greek language typesetting with \LaTeX{} and \LuaLaTeX{}. In particular, it provides support for modern Greek (either monotonic or polytonic) and ancient Greek.

1 Introduction
The \texttt{xgreek} package provides rudimentary support for Greek language typesetting with \LaTeX{} and \LuaLaTeX{}. Users will be able to typeset documents in either modern Greek (monotonic or polytonic) or ancient Greek by selecting the appropriate package option. The default “language” is monotonic Greek.

Support for \LuaLaTeX{} was provided by Javier Bezos.

2 The Source Code
According to the Unicode standard
\url{http://www.unicode.org/Public/UNIDATA/UnicodeData.txt}
the uppercase form of \texttt{GREEK SMALL LETTER EPSILON WITH TONOS} is \texttt{GREEK CAPITAL LETTER ETA WITH TONOS}. This is certainly wrong. The main reason is that accents are not part of the letter as for example is the case with \texttt{LATIN SMALL LETTER K WITH CARON}. Since, \LaTeX{} blindly follows the Unicode standard, commands like \texttt{\MakeUppercase} produce wrong output. For this reason I first need to set up the correct \texttt{uccodes} and \texttt{lccodes}.

1 \texttt{\langle xgreek \rangle}
2 \texttt{\message{Package `xgreek' version 3.0.1 by Apostolos Syropoulos}}
3 \texttt{\global\lccode"0370="0371 \global\uccode"0370="0370}
4 \texttt{\global\lccode"0371="0371 \global\uccode"0371="0370}
55 \global\lccode"03B2="03B2 \global\uccode"03B2="0392
56 \global\lccode"03B3="03B3 \global\uccode"03B3="0393
57 \global\lccode"03B4="03B4 \global\uccode"03B4="0394
58 \global\lccode"03B5="03B5 \global\uccode"03B5="0395
59 \global\lccode"03B6="03B6 \global\uccode"03B6="0396
60 \global\lccode"03B7="03B7 \global\uccode"03B7="0397
61 \global\lccode"03B8="03B8 \global\uccode"03B8="0398
62 \global\lccode"03B9="03B9 \global\uccode"03B9="0399
63 \global\lccode"03BA="03BA \global\uccode"03BA="039A
64 \global\lccode"03BB="03BB \global\uccode"03BB="039B
65 \global\lccode"03BC="03BC \global\uccode"03BC="039C
66 \global\lccode"03BD="03BD \global\uccode"03BD="039D
67 \global\lccode"03BE="03BE \global\uccode"03BE="039E
68 \global\lccode"03BF="03BF \global\uccode"03BF="039F
69 \global\lccode"03C0="03C0 \global\uccode"03C0="03A0
70 \global\lccode"03C1="03C1 \global\uccode"03C1="03A1
71 \global\lccode"03C2="03C2 \global\uccode"03C2="03A2
72 \global\lccode"03C3="03C3 \global\uccode"03C3="03A3
73 \global\lccode"03C4="03C4 \global\uccode"03C4="03A4
74 \global\lccode"03C5="03C5 \global\uccode"03C5="03A5
75 \global\lccode"03C6="03C6 \global\uccode"03C6="03A6
76 \global\lccode"03C7="03C7 \global\uccode"03C7="03A7
77 \global\lccode"03C8="03C8 \global\uccode"03C8="03A8
78 \global\lccode"03C9="03C9 \global\uccode"03C9="03A9
79 \global\lccode"03CA="03CA \global\uccode"03CA="03AA
80 \global\lccode"03CB="03CB \global\uccode"03CB="03AB
81 \global\lccode"03CC="03CC \global\uccode"03CC="03AF
82 \global\lccode"03CD="03CD \global\uccode"03CD="03A5
83 \global\lccode"03CE="03CE \global\uccode"03CE="03A9
84 \global\lccode"03D0="03D0 \global\uccode"03D0="0392
85 \global\lccode"03D1="03D1 \global\uccode"03D1="0398
86 \global\lccode"03D2="03D2 \global\uccode"03D2="03A5
87 \global\lccode"03D3="03D3 \global\uccode"03D3="03A5
88 \global\lccode"03D4="03D4 \global\uccode"03D4="03A5
89 \global\lccode"03D5="03D5 \global\uccode"03D5="03A5
90 \global\lccode"03D6="03D6 \global\uccode"03D6="03A5
91 \global\lccode"03D7="03D7 \global\uccode"03D7="03A5
92 \global\lccode"03D8="03D8 \global\uccode"03D8="03A5
93 \global\lccode"03D9="03D9 \global\uccode"03D9="03A5
94 \global\lccode"03DA="03DA \global\uccode"03DA="03A5
95 \global\lccode"03DB="03DB \global\uccode"03DB="03A5
96 \global\lccode"03DC="03DC \global\uccode"03DC="03A5
97 \global\lccode"03DD="03DD \global\uccode"03DD="03A5
98 \global\lccode"03DE="03DE \global\uccode"03DE="03A5
99 \global\lccode"03DF="03DF \global\uccode"03DF="03A5
100 \global\lccode"03E0="03E0 \global\uccode"03E0="03A5
101 \global\lccode"03E1="03E1 \global\uccode"03E1="03A5
102 \global\lccode"03E2="03E2 \global\uccode"03E2="03A5
103 \global\lccode"03E3="03E3 \global\uccode"03E3="03A5
104 \global\lccode"03E4="03E4 \global\uccode"03E4="03A5
105 \global\lccode"03E5="03E5 \global\uccode"03E5="03A5
106 \global\lccode"03E6="03E6 \global\uccode"03E6="03A5
107 \global\lccode"03E7="03E7 \global\uccode"03E7="03A5
108 \global\lccode"03E8="03E8 \global\uccode"03E8="03A5
109 \global\lccode"03E9="03E9 \global\uccode"03E9="03A5
110 \global\lccode"03EA="03EA \global\uccode"03EA="03A5
111 \global\lccode"03EB="03EB \global\uccode"03EB="03A5
112 \global\lccode"03EC="03EC \global\uccode"03EC="03A5
113 \global\lccode"03ED="03ED \global\uccode"03ED="03A5
114 \global\lccode"03EE="03EE \global\uccode"03EE="03A5
115 \global\lccode"03EF="03EF \global\uccode"03EF="03A5
116 \global\lccode"1F00="1F00 \global\uccode"1F00="0391
117 \global\lccode"1F01="1F01 \global\uccode"1F01="0391
Next I define the various strings that correspond to the standard \LaTeX\ captions. I first define the strings for monotonic Greek.

\begin{verbatim}
321 \def\prefacename{Πρόλογος}\
322 \def\refname{Αναφορές}\
323 \def\abstractname{Περίληψη}\
324 \def\bibname{Βιβλιογραφία}\
325 \def\chaptername{Κεφάλαιο}\
326 \def\appendixname{Παράρτημα}\
327 \def\contentsname{Περιεχόμενα}\
328 \def\listfigurename{Κατάλογος σχημάτων}\
329 \def\listtablename{Κατάλογος πινάκων}\
330 \def\indexname{Ευρετήριο}\
331 \def\figurename{Σχήμα}\
332 \def\tablename{Πίνακας}\
333 \def\partname{Μέρος}\
334 \def\enclname{Συνημμένα}\
335 \def\ccname{Κοινοποίηση}\
336 \def\headtoname{Προς}\
337 \def\pagename{Σελίδα}\
338 \def\seename{βλέπε}\
339 \def\alsoname{βλέπε επίσης}\
340 \def\proofname{Απόδειξη}\
341 \def\glossaryname{Γλωσσάρι}\
\end{verbatim}

Macro \texttt{polytonicnames} is invoked when polytonic Greek is the main language of the document.

\begin{verbatim}
342 \def\polytonicnames{\def\refname{Ἀναφορές}\
343 \def\indexname{Εὑρετήριο}\
344 \def\figurename{Σχῆμα}\
345 \def\headtoname{Πρὸς}\
346 \def\alsoname{βλέπε ἐπίσης}\
347 \def\proofname{Ἀπόδειξη}\
348 \def\glossaryname{Γλωσσάρι}}\
\end{verbatim}

8
Macro \texttt{\textbackslash ancient\textbackslash names} is invoked when ancient Greek is the main language of the document.

I redefine \texttt{\textbackslash today} so as to produce dates in Greek. The names of months are defined by the macro \texttt{\textbackslash gr\textbackslash month}.

Next, I define a few macros that allow one to access characters that are not usually easily accessible from the keyboard (e.g., the sampi or the koppa symbol). The list includes a command for the unicode symbol GREEK ANO TELEIA, which, in some systems, is confused with MIDDLE DOT. The use of command \texttt{\textbackslash numer\textbackslash lsign} will be explained later.
Many users prefer the use of the letters sigma and tau instead of the stigma symbol in Greek numerals, therefore, by default the \stigma command expands to “στ".

The following commands take care of the basic rules of typography. Note that the first command changes the way space is added after punctuation symbols and the last two commands force \LaTeX to add indentation space to the first paragraph after a header. Since a number of users need, for their own reasons, to be able to disable this particular feature I have introduced a new package option, namely noindentfirst, which restores the default behavior. In order to be able to do this I need the original value of the boolean variable \@afterindentfalse.

\frenchspacing
\let\@saveafterindentfalse\@afterindentfalse
\let\@afterindentfalse\@afterindenttrue
\@afterindenttrue

\ifx\directlua\undefined
non \LaTeX code
\else
Lua\LaTeX code
\fi

\ifx\directlua\undefined\else\RequirePackage{luahyphenrules}\fi

The code that follows specifies which hyphenation patterns will be active. The \LaTeX code is quite standard and depends on the babel pattern loading mechanism, while the Lua\LaTeX code uses the \HyphenRules macro, which has essentially the functionality of the \selectlanguage macro.

\DeclareOption{monogreek}{%
  \ifdef\directlua\undefined
  \language\l@monogreek
  \else
  \HyphenRules{monogreek}
  \fi%
}

\DeclareOption{polygreek}{%
  \ifdef\directlua\undefined
  \language\l@polygreek
  \else
  \HyphenRules{polygreek}
  \fi%
  \polytonicn@mes%
  \let\gr@month\gr@c@month%
}

\DeclareOption{ancientgreek}{%
  \ifdef\directlua\undefined
  \language\l@ancientgreek
  \else
  \HyphenRules{ancientgreek}
  \fi%
  \@ancientn@mes%
}
If a user wants to use the stigma symbol in Greek numerals, she should use the \texttt{stigma} option.

\begin{verbatim}
\DeclareOption{stigma}{%
  \def\stigma{\char"03DB\relax}
}
\end{verbatim}

As noted above, the new option \texttt{noindentfirst} restores the default \LaTeX{} behavior of adding no indentation to the first paragraph after any header.

\begin{verbatim}
\DeclareOption{noindentfirst}{%
  \let\@afterindentfalse\@saveafterindentfalse
}
\end{verbatim}

Nowadays it is customary in Greece to use Greek numerals without the GREEK NUMERAL SIGN at the end of numeral. Thus, the \texttt{nonumeralsign} option disables the typesetting of the GREEK NUMERAL SIGN at the end of Greek numerals.

\begin{verbatim}
\DeclareOption{nonumeralsign}{%
  \let\numer@lsign\relax
}
\end{verbatim}

By default the \texttt{monogreek} option is activated.

\begin{verbatim}
\ExecuteOptions{monogreek}
\ProcessOptions
\end{verbatim}

Now I am going to define the macros that typeset alphabetic Greek numerals. The code is borrowed from the greek option for the babel package.

\begin{verbatim}
\gr@ill@value
\end{verbatim}

When the argument of \texttt{\greeknumeral} has a value outside of the acceptable bounds \((0 < x < 999999)\) a warning will be issued (and nothing will be printed).

\begin{verbatim}
\def\gr@ill@value#1{%
\PackageWarning{xgreek}{Illegal value (#1) for greeknumeral}}
\end{verbatim}

\begin{verbatim}
\anw@false
\anw@true
\anw@print
\end{verbatim}

When a large number with three \textit{trailing} zeros is to be printed those zeros \textit{and} the numeric mark need to be discarded. As each ‘digit’ is processed by a separate macro \textit{and} because the processing needs to be expandable we need some helper macros that help remember to \textit{not} print the numeric mark (\texttt{\numer@lsign}).

The command \texttt{\anw@false} switches the printing of the numeric mark off by making \texttt{\anw@print} expand to nothing. The command \texttt{\anw@true} (re)enables the printing of the numeric mark. These macro’s need to be robust in order to prevent improper expansion during writing to files or during \texttt{\uppercase}.

\begin{verbatim}
\DeclareRobustCommand\anw@false{%
  \DeclareRobustCommand\anw@print{}%}
\end{verbatim}

\begin{verbatim}
\DeclareRobustCommand\anw@true{%
  \DeclareRobustCommand\anw@print{\numer@lsign}}
\end{verbatim}

\begin{verbatim}
\anw@true
\end{verbatim}
This command is used to typeset Greek numerals. The command uses \numer@lsign to typeset the NUMERAL SIGN. Obviously, when the user has specified the no-numeralsign option, then numeral comes out without the trailing NUMERAL SIGN. However, when a user wants to typeset a Greek numeral, the numeral must come out correctly, regardless of what appears in headers, etc. And that is exactly the reason why this command is inaccessible to users. The command \@greeknumeral needs to be fully expandable in order to get the right information in auxiliary files. Therefore we use a big \if-construction to check the value of the argument and start the parsing at the right level.

If the value is negative or zero nothing is printed and a warning is issued.

The available shorthands for 1,000 (\@m) and 10,000 (\@M) are used to save a few tokens.

If the value is too large, nothing is printed and a warning is issued.

What is left to make complete the definition of command \greeknumeral is a set of macros to produce the various digits.

As there is no “digit” representing 0 in this system, the zeros are simply discarded. When there is a large number with three trailing zeros also the numeric mark is discarded. Therefore these macros need to pass the information to each other about the (non-)translation of a zero.
The first three “digits” always have the numeric mark, except when one is discarded because it’s value is zero.

The alphabetic numbering system is not the only numbering system employed by Greeks. In fact, Greeks used various systems that are now known as acrophonic numbering systems. Many scholars are familiar with the acrophonic Attic
numbering system and the the command \atticnum can be used to generate acro-
phonic Attic numerals. The acrophonic Attic numbering system, like the Roman
one, employs letters to denote important numbers. Multiple occurrence of a let-
ter denote a multiple of the “important” number, e.g., the letter I denotes 1, so
III denotes 3. Here are the basic digits used in the acrophonic Attic numbering
system:

- I denotes the number one (1)
- II denotes the number five (5)
- Δ denotes the number ten (10)
- Η denotes the number one hundred (100)
- Χ denotes the number one thousand (1000)
- Μ denotes the number ten thousands (10000)

Moreover, the letters Δ, Η, Χ, and Μ under the letter Π (a form of Π) denote five
times their original value. In particular, the symbol Ρ denotes the number 50, the
symbol Ρ denotes the number 500, the symbol Ρ denotes the number 5000, and
the symbol Ρ denotes the number 50,000. It must be noted that the numbering
system does not provide negative numerals or a symbol for zero.

Now, let me definite the macro \@atticnum. This macro uses one integer variable
(or counter in \TeX's jargon.)
\newcount\@attic@num

The macro \@atticnum is also defined as a robust command.
\DeclareRobustCommand*{\@attic@num}[1]{%
After assigning to variable \@attic@num the value of the macro’s argument, we
make sure that the argument is in the expected range, i.e., it is greater than zero,
and less or equal to 249999. In case it isn’t, it simply produces a \space, warns the
user about it and quits. Although, the \atticnum macro is capable to produce an
Athenian numeral for even greater intergers, the following argument by Claudio
Beccari convised me to place this upper limit:

According to psychological perception studies (that ancient Atheni-
ans and Romans perfectly knew without needing to study Freud and
Jung) living beings (which includes at least all vertebrates, not only
humans) can perceive up to four randomly set objects of the same kind
without the need of counting, the latter activity being a specific ac-
quired ability of human kind; the biquinary numbering notation used
by the Athenians and the Romans exploits this natural characteristic
of human beings.
\@attic@num#1\relax
\ifnum\@attic@num<\@ne%
\space%
\else
\space%
Having done all the necessary checks, it is possible to proceed with the actual computation. If the number is greater than 49999, then it certainly has at least one 𐅇 “digit”. The macro finds all such digits by continuously subtracting 50000 from \@attic@num, until \@attic@num becomes less than 50000.

Next the macro checks for tens of thousands.

Since a number can have only one 𐅆 “digit” (equivalent to 5000), it is easy to check whether is should have one and produce the corresponding numeral when it does have one.

The macro should also check for thousands, the same way it checked for tens of thousands.

Since a number can have at most one 𐅅 “digit” (equivalent to 500), this should be handled the way the macro handled the case of the five thousands “digit”.

It is time to check hundreds, which follow the same pattern as thousands.

A numeral can have only one 𐅄 “digit” (equivalent to 50).

The macro now checks now for tens digit.

Finally, it has to check for fives and the digits 1, 2, 3, and 4.
The command \atticnum has one argument, which is a counter. It calls the command \@@atticnum to process the value of the counter.
\begin{verbatim}
def\atticnum#1{\expandafter\@@atticnum\expandafter{\the#1}}
\end{verbatim}

The command \atticnum is a wrapper that declares a new counter in a local scope, assigns to it the value of the argument of the command and calls the macro \atticnum. This way the command can process correctly either a number or a counter.
\begin{verbatim}
def\atticnum#1{\@attic@num#1\relax\@atticnum{\@attic@num}}
\end{verbatim}

Here I redefine the macros \alph and \Alph. First, I define some placeholders
\begin{verbatim}
\let\latin@alph\@alph
\let\latin@Alph\@Alph
\end{verbatim}

Then I define the Greek versions; the additional \expandafter s are needed in order to make sure the table of contents will be correct (e.g., when there are appendices).
\begin{verbatim}
def\greek@alph#1{\expandafter\@greeknumeral\expandafter{\the#1}}
def\greek@Alph#1{\expandafter\@Greeknumeral\expandafter{\the#1}}
\end{verbatim}

By default Greek alphabetic enumerals instead of Latin numerals are used to enumerate items in an enumeration environment.
\begin{verbatim}
\let\@alph\greek@alph
\let\@Alph\greek@Alph
\end{verbatim}

If for some reason, one needs to have the Latin numerals back, then she has to invoke command \nogreekalph. And if she wants to switch back, then she has to use the \greekalph command:
\begin{verbatim}
def\nogreekalph{\let\@alph\latin@alph
\let\@Alph\latin@Alph}
def\greekalph{\let\@alph\greek@alph
\let\@Alph\greek@Alph}
\end{verbatim}

We provide the \setlanguage command which activates the hyphenation patterns of some other language. It is similar to babel's \selectlanguage, but we opted to use a new name to avoid possible name conflicts. Valid arguments include monogreek, polygreek, ancientgreek, and american. As was noted previously, package luahyphenrules provides the command \HyphenRules which has exactly the same functionality as this command. So when using Lua\LaTeX users will actually use the \HyphenRules command.
The macros \grtoday and \Grtoday produce the current date, only that the month and the day are shown as greek numerals instead of arabic as it is usually the case. In addition, the two commands differ in that the later produces the Greek numerals in uppercase.