The Serbian Cyrillic Language in the babel system

Uroš Stefanović*

September 30, 2022

Serbian Cyrillic Language

The file *serbanc.dtx* defines all the language definition macros for Serbian language, typeset in a Cyrillic script.

For this language the character " is made active. In table 1 an overview is given of its purpose. One of the reasons for this is that, in the Serbian language, some special characters are used.

"- An explicit hyphen sign, allowing hyphenation in the rest of the word; inserts a hyphen which is repeated at the beginning of the next line (recommended to use for compound words with hyphen).

"| Enables ligature at this position.

"" Similar to "- but prints no hyphen sign.

"~ Compound word mark without a breakpoint, prints hyphen prohibiting hyphenation at the point.

"=" A compound word mark with a breakpoint, prints hyphen allowing hyphenation in the composing words.

"' German opening double quote (looks like „). 

"' German closing double quote (looks like “). 

"'(if the quotes attribute is used) Closing double quote (looks like ”).

"< French opening double quote (looks like <<).

"> French closing double quote (looks like >>).

Table 1: The extra definitions made by *serbanc.ldf*

Macro \today prints the date in Serbian. Alternatively, if attribute datei is used, \today prints the current date, but prints ‘jyun’ and ‘jyul’ for ‘June’ and ‘July’. If you prefer to use ‘jyun’ and ‘jyul’ instead of default ‘jun’ and ‘jul’, use the datei attribute. Also, the \today* macro prints the date without dot after the year (used when after the date is the punctuation mark, such as comma). Alternatively, the commands \todayRoman and \todayRoman* prints the current date using Roman numerals for months; \todayGen and \todayGen* prints the current month name in the genitive case, and \todayArabic and \todayArabic* prints the current month as a number.

The alphabetical enumerations in texts use the Cyrillic alphabet and alphabetic order (all 30 letters of the Serbian language are used). Also, the Serbian language allows enumeration with the Latin alphabet. If the Latin alphabet is used in the enumeration the letters q, w, x and y are omitted by the rules of the Serbian language (22 letters are used in that case). However, if the user wants to use the English alphabet for the enumeration (26 letters), this option is also available. One can manually switch the enumeration alphabet with the commands \enumCyr, \enumLat and \enumEng. This commands can be used after the \begin{document} when the *serbanc* language is active. In

*urostajms@gmail.com*
principle, enumerations are a matter for class and style designers but the same can be said also about
things, other than enumerations, such as names of sections and bibliography lists.

Apart from defining shorthands, we need to make sure that the first paragraph of each section is
indented. Furthermore, the following new math operators are defined: \sh, \ch, \tg, \ctg, \arctg, \\arcctg, \th, \cth, \arsh, \arch, \arth, \cosesc, \Prob, \Expect, \Variance, \arcsec, \\arcosec, \sech, \cosech, \arsech, \arcosech, \NZD, \nzd, \NZS, \nzs. Cyrillic letters in math
mode can be typed with the aid of text commands such as \textbf, \textsf, \textit, \texttt, e.t.c.

By default, ekavian spelling is enabled. Ijekavian spelling can be enabled by setting the attribute
to \ije. To set an attribute, put the \texttt{languageattribute} macro within a document preamble after
\texttt{babel}, for example,

\texttt{\usepackage[english,serbian]{babel}}
\texttt{\languageattribute{serbian}{ije}}

Setting the \ije attribute changes the built-in strings (caption names). For example, the part will
be entitled as ‘Део’ by default and as ‘Дио’ if the Serbian language attribute is set to \ije. Same
result can be achieved using a modifier as follows:

\texttt{\usepackage[english,serbian.ije]{babel}}

Using a modifier in a package option is often better. A modifier is set after the language name, and
is prefixed with a dot (only when the language is set as package option — neither global options nor
the main key accept them). Also, it’s possible to use more than one attribute:

\texttt{\usepackage[english,serbian.ije.datei.quotes]{babel}}

The file \texttt{serbianc.1df} is designed to work both with legacy non-unicode (8-bit) and new Unicode
encodings of the source document files (input encodings) and of the font files (font encodings). This
is achieved by excluding (bypassing) the \texttt{\cyr...} macros, which map every letter in a source file with
given input encoding to a corresponding code point in a font file with a given font encoding when
running modern engines, such as \texttt{LuaLATEX} or \texttt{XeLATEX}, in native Unicode mode instead of legacy
engines, such as \texttt{LATEX} or \texttt{PDFLATEX}, or Unicode engines in a compatibility (8-bit) mode.

For \texttt{LuaLATEX} or \texttt{XeLATEX} one needs to load the \texttt{fontspec} package. The following example shows
how to load Computer Modern Unicode (CMU) fonts (which is a part of all modern \texttt{LATEX}
distributions), and also to get correct italic shape of the letters \texttt{б}, \texttt{г}, \texttt{д}, \texttt{п} and \texttt{т} for Serbian language:

\texttt{\usepackage[fontspec]}
\texttt{\defaultfontfeatures{Ligatures={TeX},Language=Serbian,Script=Cyrillic}}
\texttt{\setmainfont{CMU Serif}}
\texttt{\setsansfont{CMU Sans Serif}}
\texttt{\setmonofont{CMU Typewriter Text}}
\texttt{\usepackage[english,serbian]{babel}}

The code

The macro \texttt{\LdfInit} takes care of preventing this file from loading more than once, checking the
category code of the \texttt{@} sign, etc.

1 ⟨∗code⟩
2 \texttt{\LdfInit{serbianc}{captionsserbianc}}

First, we check if \texttt{LuaLATEX} or \texttt{XeLATEX} is running. If it is, we set the boolean key
\texttt{\if@srbc@uni@ode} to true. It will be used to eliminate \texttt{\cyr...} commands, which were introduced
in \texttt{LATEX} to handle various Cyrillic input encodings. With the introduction of Unicode, \texttt{LATEX}
is moving to universal input encoding, so we consider these \texttt{\cyr...} commands obsolete. However, they
are still preserved for backward compatibility in case \texttt{LATEX} or \texttt{PDFLATEX} are running.

3 \texttt{\ifdefined\if@srbc@uni@ode}
4 \texttt{\PackageError{babel}{\if@srbc@uni@ode already defined.}}
Check if hyphenation patterns for the Serbian language have been loaded in language.dat. Namely, we check for the existence of \l@serbianc. If it is not defined, we declare Serbian as dialect for the default language number 0, which is almost certainly English.

There is a limited list of encodings appropriate for Serbian Cyrillic text. We will look at which one of them is declared and keep its name in the macro \cyrillicencoding. The correct 7-bit Cyrillic encoding is OT2. The correct 8-bit Cyrillic encodings are T2A (default for 8-bit compilers) and X2. The correct utf8 encodings are TU (default for \TeX and Lua\TeX), EU1 (obsolete, formerly used for \TeX), EU2 (obsolete, formerly used for Lua\TeX).

In 8-bit (\LaTeX) mode, the user may choose a different non-unicode Cyrillic encoding—X2 or OT2. If one wants to use another font encoding rather than default (T2A), he or she has to load the corresponding file before babel.sty.

Remember that, for the Serbian language, the T2A encoding is better than X2, because X2 does not contain Latin letters, and users should pay attention and switch the language every time they want to type a Latin word inside a Serbian phrase or vice versa.

We parse the \cdp@list containing encodings known to \LaTeX in the order in which they have been loaded by the time babel is called. We set the \cyrillicencoding to the last loaded encoding in the list of supported Cyrillic encodings: OT2, X2, T2A. In Unicode mode, \cyrillicencoding is set to TU by \fontspec. Nevertheless, here we provide similar definitions; 8-bit encodings are kept for Unicode compilers (Lua\LaTeX and Xe\LaTeX) since they can run in compatibility (8-bit) mode.

\def\@setcyrillicencoding{%
  \def\sce@a##1##2{%
    \edef\sce@b{##1}%
    \edef\sce@c{##2}%
    \ifx\sce@b\sce@c%
      \let\cyrillicencoding\sce@c%
    \fi}%
  \def\cdp@elt##1##2##3##4{%
    \sce@a{##1}{OT2}%
    \sce@a{##1}{X2}%
    \sce@a{##1}{T2A}%
    \if@srbc@uni@ode%
      \sce@a{##1}{TU}%
    \fi}%
  \cdp@list
}
We define the macro \Serbianc simply as an alias for \selectlanguage{serbianc}.

We define \cyrillictext and its alias \cyr; these macros are intended for use within the babel macros and do not perform the complete change of the language.

In particular, they do not change the captions and the name of current language stored in the macro \languagename. This inconsistency might break some assumptions embedded into babel. For example, the \iflanguage macro will fail.

Furthermore, \cyrillictext does not activate shorthands, so "<", ",", ",", e.t.c. will not work. Lastly, \cyrillictext does not write its trace to .aux file, which might result in wrong typesetting of table of contents, list of tables and list of figures in multilingual documents.

For these reasons, the use of the declaration \cyrillictext and its aliases in ordinary text is strongly discouraged. Instead of the declaration \cyrillictext, it is recommended to use \Serbianc or the command \foreignlanguage defined in the babel core; their functionality is similar to \selectlanguage{serbianc}, but they do not change caption names, dates and shorthands.

The macro \textcyrillic takes an argument which is then typeset using the \cyrillictext declaration.

For Serbian, the " character is made active. This is done once; later on, its definition may vary. Other languages in the same document may also use the " character for shorthands; we specify that the Serbian group of shorthands should be used. We save the original double quote character in the \dq macro to keep it available. The shorthand "- should be used in places where a word contains an explicit hyphenation character. According to the rules of the Serbian language, when a word break occurs at an explicit hyphen, it must appear both at the end of the first line and at the beginning of the second line.
The \cyrdash macro will be defined if it hadn’t already been defined in a fontenc file. For T2A and X2 fonts, \cyrdash will be placed in the code of the English emdash.

\providecommand{\cyrdash}{\hbox to.8em{--\hss--}}

The macro \captionsserbianc defines all strings used in the four standard documentclasses provided by \LaTeXX.

\if@srbc@uni@ode
\addto\captionsserbianc{%
  \def\prefacename{ Предговор }%
  \def\refname{ Литература }%
  \def\abstractname{ Сажетак }%
  \def\bibname{ Библиографиjа }%
  \def\chaptername{ Глава }%
  \def\appendixname{ Додатак }%
  \def\contentsname{ Садржаj }%
  \def\listfigurename{ Списак слика }%
  \def\listtablename{ Списак табела }%
  \def\indexname{ Индекс }%
  \def\figurename{ Слика }%
  \def\tablename{ Табела }%
  \def\partname{ Дио }%
  \def\enclname{ Прилози }%
  \def\ccname{ Копиjе }%
  \def\headtoname{ Прима }%
  \def\pagename{ страна }%
  \def\seename{ види }%
  \def\alsoname{ види такође }%
  \def\proofname{ Доказ }%
  \def\glossaryname{ Речник }%
}%
\addto\captionsserbianc@ijekav{%
  \def\partname{ Део }%
  \def\glossaryname{ Рjечник }%
}{}
\else
\addto\captionsserbianc{%
  \def\prefacename{{\cyrd\cyre\cyrd\cyrg\cyro\cyrv\cyro\cyrr}}%
  \def\refname{{\cyrl\cyri\cyrt\cyre\cyrl\cyra}}%
  \def\abstractname{{\cyr\cyrzh\cyre\cyrt\cyra}}%
  \def\bibname{{\cyrb\cyri\cyrb\cyrl\cyri\cyro\cyrg\cyrr\cyra}}%
  \def\chaptername{{\cyrg\cyrl\cyra\cyrv\cyra}}%
  \def\appendixname{{\cyrd\cyro\cyrd\cyra\cyrt\cyra\cyrk}}%
  \def\contentsname{{\cyr\cyrd\cyrr\cyrzh\cyra\cyrje}}%
  \def\listfigurename{{\cyr\cyrp\cyri\cyrs\cyra\cyrs\cyrl\cyri\cyrk\cyra}}%
  \def\listtablename{{\cyr\cyrp\cyri\cyrs\cyra}}%
  \def\indexname{{\cyr\cyrn\cyrd\cyre\cyrk}}%
  \def\figurename{{\cyr\cyrl\cyri\cyrk\cyra}}%
  \def\tablename{{\cyrt\cyra\cyrb\cyre\cyrl\cyra}}%
  \def\partname{{\cyr\cyrd\cyre\cyro}}%
  \def\enclname{{\cyr\cyrr\cyri\cyrl\cyro\cyrz\cyri}}%
  \def\ccname{{\cyr\cyrk\cyro\cyrp\cyri}}%
  \def\headtoname{{\cyrr\cyri\cyrd\cyri}}%
  \def\pagename{{\cyr\cyrs\cyri\cyrd\cyri}}%
  \def\seename{{\cyr\cyrv\cyri\cyrd\cyri}}%
  \def\alsoname{{\cyr\cyrv\cyri\cyrd\cyri\cyrt\cyra\cyrk\cyri}}%
  \def\proofname{{\cyrd\cyro\cyrz}}%
  \def\glossaryname{{\cyr\cyrch\cyrn\cyri\cyrk}}%
}{}
\let\captionsserbianc@ijekav=\captionsserbianc
\addto\captionsserbianc@ijekav{%
  \def\partname{Дио}%
  \def\glossaryname{Рјечник}%
}{}
\else
\fi
The macro \dateserbianc redefines the commands \today, \today*, \todayRoman and \todayRoman* to produce Serbian dates. Also, the commands \todayGen, \todayGen*, \todayArabic and \todayArabic* are provided.
The Serbian hyphenation patterns can be used with \lefthyphenmin and \righthyphenmin set to 2. (Actually, the “official” definition allows even one character for \left hyphen, but it is recommended to use the value two for the better results.)

\providehyphenmins{\CurrentOption}{\tw@\tw@}
\providehyphenmins{serbianc}{\tw@\tw@}

We instruct babel to switch font encoding using earlier defined macros \cyrillictext and \latintext.

\addto\extrasserbianc{\cyrillictext}
\addto\noextrasserbianc{\latintext}

Also, we specify that the Serbian group of shorthands should be used.

\addto\extrasserbianc{\languageshorthands{serbianc}}
\addto\extrasserbianc{\bbl@activate{"}}
\addto\extrasserbianc{\bbl@deactivate{"}}

Serbian typesetting requires \frenchspacing. So, we add commands to \extrasserbianc and \noextrasserbianc to turn it on and off, respectively.

\addto\extrasserbianc{\bbl@frenchspacing}
\addto\noextrasserbianc{\bbl@nonfrenchspacing}

In Serbian, the first paragraph of each section should be indented.

\let@afterORI\@afterindentfalse
\def\bbl@serbiancindent{\let@afterindentfalse\@afterindenttrue\@afterindenttrue}
\def\bbl@nonserbiancindent{\let@afterindentfalse\@afterindenttrue\@afterindentfalse}
\addto\extrasserbianc{\bbl@serbiancindent}
\addto\noextrasserbianc{\bbl@nonserbiancindent}

We redefine the macro \Alph, which now produces (uppercase) Cyrillic letters instead of Latin ones when Serbian is switched on. Also we will define Serbian Latin and English alphabets so the user can choose which alphabet to use through the commands \enumCyr, \enumLat and \enumEng (or even to switch from one enumeration to another).

\newcount\srbc@lettering \srbc@lettering=\z@
\addto\extrasserbianc{\babel@save\@Alph \let@Alph@srbc@Alph}

7
An iijkev attribute changes the default behavior and activates an alternative set of captions suitable for typesetting in iijkevian dialect. The quotes attribute changes the "'" shorthand to produce " as closing quotation mark, instead of the traditional " quotation mark of Serbian language. Also, the datei attribute will produce 'jünü' and 'jyüli' instead 'jünü' and 'julyi' for date.
Some math functions in Serbian math books have other names: e.g. \texttt{\sinh} in Serbian is written as \texttt{sh} etc. So we define a number of new math operators.

\begin{verbatim}
\def\sh{\mathop{\operator@font sh}\nolimits}
\def\ch{\mathop{\operator@font ch}\nolimits}
\def\tg{\mathop{\operator@font tg}\nolimits}
\def\ctg{\mathop{\operator@font ctg}\nolimits}
\def\arctg{\mathop{\operator@font arctg}\nolimits}
\def\arcctg{\mathop{\operator@font arcctg}\nolimits}
\addto\extrasserbianc{\babel@save\th}
\let\ltx@th\th
\def\th{\textormath{\ltx@th}{\mathop{\operator@font th}\nolimits}}
\addto\extrasserbianc{\babel@save\nzs \babel@save\nzd}
\if@srbc@uni@ode
\def\nzs{\mathop{\mathrm{нзс}}\nolimits}
\def\nzd{\mathop{\mathrm{нзд}}\nolimits}
\def\NZS{\mathop{\mathrm{НЗС}}\nolimits}
\def\NZD{\mathop{\mathrm{НЗД}}\nolimits}
\else
\def\nzs{\mathop{\textnormal{нзс}}\nolimits}
\def\nzd{\mathop{\textnormal{нзд}}\nolimits}
\def\NZS{\mathop{\textnormal{НЗС}}\nolimits}
\def\NZD{\mathop{\textnormal{НЗД}}\nolimits}
\fi
\end{verbatim}

The macro \texttt{\ldf@finish} takes care of looking for a configuration file, setting the main language to be switched on at \texttt{\begin{document}} and resetting the category code of \texttt{@} to its original value.