

Estonian language support for **babel**

Enn Saar*, Jaan Vajakas†

2015/08/05, v1.1a

The file `estonian.dtx` defines the language definition macro's for the Estonian language.

This file was written as part of the TWGML project, and borrows heavily from the `babel` German and Spanish language files `germanb.ldf` and `spanish.ldf`.

Estonian has the same umlauts as German (ä, ö, ü), but in addition to this, we have also õ, and two recent characters š and ž, so we need at least two active characters. We shall use " and ~ to type Estonian accents on ASCII keyboards (in the 7-bit character world). Their use is given in table 1. These active accent

~o	\~o, (and uppercase);
"a	\"a, (and uppercase);
"o	\"o, (and uppercase);
"u	\"u, (and uppercase);
~s	\v s, (and uppercase);
~z	\v z, (and uppercase);
"	disable ligature at this position and allow hyphenation at this position;
"-	like \-, but allowing hyphenation in the rest of the word;
"‘	for Estonian low left double quotes (same as German);
"’	for Estonian right double quotes;
"<	for French left double quotes (also rather popular)
">	for French right double quotes.

Table 1: The extra definitions made by `estonian.ldf`

characters behave according to their original definitions if not followed by one of the characters indicated in that table; the original quote character can be typed using the macro `\dq`.

*saar@aai.ee, original author (inactive)

†jaanvajakas@hotmail.ee, maintainer since version 1.0k

1 Usage guidelines

1.1 Overview and usage example

In short, it is recommended to include lines like the following in the preamble:

- in L^AT_EX:

```
\usepackage[utf8]{inputenc}
\usepackage[T1]{fontenc}
\usepackage{mathptmx} % or \usepackage{lmodern} or something else
\usepackage[estonian .notilde]{babel}
```

- in XeLaTeX or LuaLaTeX:¹

```
\usepackage[estonian .notilde]{babel}
\usepackage{fontspec}
```

When saving your file, make sure your text editor saves it in the UTF-8 encoding.

In the following subsections, the rationale of these options is explained. Some authors have also advised that ligatures should be turned off in Estonian; the last subsection tries to explain why and how.

1.2 Use the T1 font encoding and avoid CM

If using the Estonian package in L^AT_EX, it is recommended to choose the T1 output encoding (also known as the Cork encoding). It will give you better hyphenation, as the standard Estonian hyphenation file `eehyph.tex` is in this encoding. You can choose the T1 output encoding with `\usepackage[T1]{fontenc}`.

If you like Computer Modern (CM), the default font, then we recommend using its successor Latin Modern (`\usepackage{lmodern}`) instead — it is almost identical but has the tilde in the letter “ö” slightly lower, looking more natural. (In the OT1 encoding, the Estonian package takes special care to lower the tilde, but that feature is not supported for the T1 output encoding since version 1.0k of the Estonian package, as it created many issues of its own.)

In XeLaTeX and LuaLaTeX, the default font (CM) does not support accented letters and therefore the `fontspec` package (or `xltxtra`) has to be used.

1.3 Use UTF-8 and disable the tilde shorthands

In the early 1990s, handling accented letters was a problem for many programs. So shorthands for accented letters in Estonian (õ, "a etc.) were created. Nowadays the UTF-8 standard is widespread, allowing to represent letters from almost all the world’s languages in a single text file. Therefore it is recommended to use UTF-8 as input encoding and type Estonian accented characters directly, instead

¹In XeLaTeX and LuaLaTeX, you may want to use the more modern `polyglossia` package instead of `babel`, although the latter also works.

of using the shorthands. We also recommend UTF-8 over cp1257, the default encoding for Estonian on Windows, since UTF-8 contains more characters and is more likely understood by the text editors of your foreign partners.

In addition, a problem is caused by the shorthands starting with tilde. Namely, in \TeX , the `~` command originally means non-breaking space: if `babel` is not used then e. g. `U.~S.` is rendered as “U. S.”. However, with `\usepackage[estonian]{babel}` in preamble, `U.~S.` is rendered as “U.Š.” by default. If you don’t need the shorthands starting with tilde, you can disable them and restore the original behavior of `~` by using the option `notilde`, like this:

```
\usepackage[estonian .notilde]{babel}
```

The option `notilde` was introduced in version 1.1 of the Estonian package (released 2014/02/21). If you need compatibility with older versions, write instead

```
\usepackage[estonian]{babel}
\makeatletter\addto\extrastonian{\bbl@deactivate{~}}\makeatother
\addtocaptionsestonian{% Redefine captions containing ~o
  \def\abstractname{Kokkuv~ote}%
  \def\proofname{T~oestus}%
  \def\glossaryname{S~onastik}%
}
```

If you don’t disable the tilde shorthands, you must write `U.\nobreak{ } S.`

1.4 Ligatures

Using ligatures in Estonian is discouraged by some authors; it may slightly harm readability and has been uncommon in the past.² By default, \TeX creates the ligatures `fi`, `ff`, `fl`, `ffi` and `ffl` in the CM fonts (other fonts may have other ligatures). You can disable ligatures one by one (e. g. `f\|ii"-ber` or `šef"|lu"-se`³), but if you want to get rid of all ligatures, then

- in \LaTeX , include the following two lines in your preamble:⁴

```
\usepackage{microtype}
\DisableLigatures[f]{encoding = *, family = * }
```

²In their “Lühike \LaTeX iõpetus” (1994), Hans Ibrus and Enn Saar say that \TeX ’s ligatures are “not recommended” in Estonian and give the word “fiber” as a (perhaps particularly) bad example. Indeed, “ii” denotes a single long vowel, but the ligature seems to suggest that instead the first two letters “f” and “i” are grouped together. Ligatures are practically absent from Estonian books published throughout the Soviet period (e. g. mathematical books have no ligature in “definitsoon”). During Estonia’s first independence period, ligatures did occasionally appear in books: e. g. Borkvell’s “Tasapinnalise ja ruumilise analüütilise geomeetria põhijooni” from 1937 (fi ligature in “definitsoon”) or “Eesti Entsüklopeedia” from 1932–1937 (having e. g. “Affinsus” with ligature fi, but “Affiks” without ligatures and “affiinseks” with ligature ff — so ligatures were used quite randomly and arguably for the convenience of the typesetter rather than the reader).

³The commands `\|` and `"|` both disable ligature; the latter also enables hyphenation *šef-luse*. Unfortunately both disable hyphenation in the rest of the word; that’s why we need the `"-` here.

⁴Note that these two lines disable only ligatures starting with `f`. You would not want to apply `\DisableLigatures` to all letter combinations, since it also disables kerning (and letter combinations like “Ve” and “AV” look much nicer with kerning).

- in XeLaTeX and LuaLaTeX, use this line instead (replace the font name):

```
\setmainfont[Ligatures={NoRequired,NoCommon,NoContextual}]{Font Name}
```

2 Implementation

The macro `\LdfInit` takes care of preventing that this file is loaded more than once, checking the category code of the `@` sign, etc.

```
1 <*code>
2 \LdfInit{estonian}\captionsestonian
```

If Estonian is not included in the format file (does not have hyphenation patterns), we shall use English hyphenation.

```
3 \ifx\l@estonian\@undefined
4 \nopatterns{Estonian}
5 \addialect\l@estonian0
6 \fi
```

Now come the commands to switch to (and from) Estonian.

`\captionsestonian` The macro `\captionsestonian` defines all strings used in the four standard documentclasses provided with L^AT_EX.

```
7 \addto\captionsestonian{%
8 \def\prefacename{Sissejuhatus}%
9 \def\refname{Viited}%
10 \def\bibname{Kirjandus}%
11 \def\appendixname{Lisa}%
12 \def\contentsname{Sisukord}%
13 \def\listfigurename{Joonised}%
14 \def\listtablename{Tabelid}%
15 \def\indexname{Indeks}%
16 \def\figurename{Joonis}%
17 \def\tablename{Tabel}%
18 \def\partname{Osa}%
19 \def\enclname{Lisa(d)}%
20 \def\ccname{Koopia(d)}%
21 \def\headtoname{}%
22 \def\pagename{Lk.}%
23 \def\seename{vt.}%
24 \def\alsoname{vt. ka}%
25 \def\abstractname{Kokkuv~ote}%
26 \def\chaptername{Peat"ukk}%
27 \def\proofname{T~oestus}%
28 \def\glossaryname{S~onastik}%
29 }
```

`\dateestonian` The macro `\dateestonian` redefines the command `\today` to produce Estonian dates.

```

30 \begingroup \catcode'\active
31 \def\x{\endgroup
32 \def\month@estonian{\ifcase\month\or
33   jaanuar\or veebruar\or m"arts\or aprill\or mai\or juuni\or
34   juuli\or august\or september\or oktoober\or november\or
35   detsember\fi}}
36 \x
37 \def\dateestonian{%
38 \def\today{\number\day.\space\month@estonian
39 \space\number\year.\space a.}}

```

Some useful macros, copied from the spanish package (and renamed `es@...` to `et@...`).

```

40 \def\et@sdef#1{\babel@save#1\def#1}
41
42 \@ifundefined{documentclass}
43 {\let\ifet@latex\iffalse}
44 {\let\ifet@latex\iftrue}

```

`\extrasestonian` The macro `\extrasestonian` will perform all the extra definitions needed for Estonian. The macro `\noextrasestonian` is used to cancel the actions of `\extrasestonian`. For Estonian, " is made active and has to be treated as 'special' (~ is active already).

```

45 \initiate@active@char{"}
46 \initiate@active@char{~}
47 \addto\extrasestonian{\languageshorthands{estonian}}
48 \addto\extrasestonian{\bbl@activate{"}\bbl@activate{~}}

```

`notilde` The option `notilde` disables the shorthands starting with ~, restoring the original function of ~ as non-breaking space.

```

49 \bbl@declare@ttribute{estonian}{notilde}{\addto\extrasestonian{\bbl@deactivate{~}}}

```

Estonian does not use extra spaces after sentences.

```

50 \addto\extrasestonian{\bbl@frenchspacing}
51 \addto\noextrasestonian{\bbl@nonfrenchspacing}

```

`\estonianhyphenmins` For Estonian, `\lefthyphenmin` and `\righthyphenmin` are both 2.

```

52 \providehyphenmins{\CurrentOption}{\tw@\tw@}

```

The standard T_EX accents are too high for Estonian typography, we have to lower them (following the babel German style). For umlauts, we can use `\umlautlow` in `babel.ldf`.

```

53 \addto\extrasestonian{\umlautlow}
54 \addto\noextrasestonian{\umlauthigh}

```

Redefine tilde (as in `spanish.ldf`). In case of L^AT_EX, we redefine the internal macro for the OT1 encoding because in case of T1, the display and hyphenation of words containing `\~o` works better without redefining it (e. g. words containing

`\et@gentilde` are not hyphenated unless `\allowhyphens` is used; when copied from Acrobat Reader, pasting an `õ` generated using `\et@gentilde{o}` gives `~o` rather than `õ`; when the `times` package is used with T1 encoding, `\et@gentilde` places the tilde through the letter o). In plain T_EX there is no encoding infrastructure, so we just redefine `\~`.

```
55 \ifet@latex
56   \addto\extrasestonian{%
57     \expandafter\et@sdef\csname OT1\string\~\endcsname{\et@gentilde}}
58 \else
59   \addto\extrasestonian{\et@sdef\~{\et@gentilde}}
60 \fi
```

`\et@gentilde`

```
61 \def\et@gentilde#1{%
62   \if#1s\v{#1}\else\if#1S\v{#1}\else%
63   \if#1z\v{#1}\else\if#1Z\v{#1}\else%
64   \et@newtilde{#1}%
65   \fi\fi\fi\fi}
```

`\et@newtilde` For a detailed explanation of the following code see the definition of `\lower@umlaut` in `babel.dtx`.

```
66 \def\et@newtilde#1{%
67   \leavevmode\bgroup\U@D 1ex%
68   {\setbox\z@\hbox{\char126}\dimen@ -.45ex\advance\dimen@\ht\z@
69     \ifdim 1ex<\dimen@ \fontdimen5\font\dimen@ \fi}%
70   \accent126\fontdimen5\font\U@D #1%
71   \egroup}
```

We save the double quote character in `\dq`, and tilde in `\til`.

```
72 \begingroup \catcode'\ "12
73 \edef\x{\endgroup
74   \def\noexpand\dq{"}
75   \def\noexpand\til{~}}
76 \x
```

If the encoding is T1, we have to tell T_EX about our redefined accents.

```
77 \ifx\fontencoding\bb1@t@one
78   \DeclareTextComposite{\~}{T1}{s}{178}
79   \DeclareTextComposite{\~}{T1}{S}{146}
80   \DeclareTextComposite{\~}{T1}{z}{186}
81   \DeclareTextComposite{\~}{T1}{Z}{154}
82   \DeclareTextComposite{\~}{T1}{'}{17}
83   \DeclareTextComposite{\~}{T1}{'}{18}
84   \DeclareTextComposite{\~}{T1}{<}{19}
85   \DeclareTextComposite{\~}{T1}{>}{20}
```

If the encoding differs from T1, we expand the accents, enabling hyphenation beyond the accent. In this case T_EX will not find all possible breaks, and we have to warn people.

```

86 \else
87 \wlog{Warning: Hyphenation would work better for the T1 encoding.}
88 \fi

    Now we define the shorthands: umlauts,
89 \declare@shorthand{estonian}{"a}{\textormath{"{a}\allowhyphens}{\ddot a}}
90 \declare@shorthand{estonian}{"A}{\textormath{"{A}\allowhyphens}{\ddot A}}
91 \declare@shorthand{estonian}{"o}{\textormath{"{o}\allowhyphens}{\ddot o}}
92 \declare@shorthand{estonian}{"O}{\textormath{"{O}\allowhyphens}{\ddot O}}
93 \declare@shorthand{estonian}{"u}{\textormath{"{u}\allowhyphens}{\ddot u}}
94 \declare@shorthand{estonian}{"U}{\textormath{"{U}\allowhyphens}{\ddot U}}

    German and French quotes,
95 \declare@shorthand{estonian}{"‘}{%
96 \textormath{\quotedblbase}{\mbox{\quotedblbase}}}
97 \declare@shorthand{estonian}{"’}{%
98 \textormath{\textquotedblleft}{\mbox{\textquotedblleft}}}
99 \declare@shorthand{estonian}{"<}{%
100 \textormath{\guillemotleft}{\mbox{\guillemotleft}}}
101 \declare@shorthand{estonian}{">}{%
102 \textormath{\guillemotright}{\mbox{\guillemotright}}}

    tildes and carons
103 \declare@shorthand{estonian}{~o}{\textormath{\~{o}\allowhyphens}{\tilde o}}
104 \declare@shorthand{estonian}{~O}{\textormath{\~{O}\allowhyphens}{\tilde O}}
105 \declare@shorthand{estonian}{~s}{\textormath{\v{s}\allowhyphens}{\check s}}
106 \declare@shorthand{estonian}{~S}{\textormath{\v{S}\allowhyphens}{\check S}}
107 \declare@shorthand{estonian}{~z}{\textormath{\v{z}\allowhyphens}{\check z}}
108 \declare@shorthand{estonian}{~Z}{\textormath{\v{Z}\allowhyphens}{\check Z}}

    and some additional commands:
109 \declare@shorthand{estonian}{"-}{\nobreak\-\bbl@allowhyphens}
110 \declare@shorthand{estonian}{"}{|}{%
111 \textormath{\nobreak\discretionary{-}{-}{\kern.03em}%
112 \allowhyphens}{}}
113 \declare@shorthand{estonian}{""}{\dq}
114 \declare@shorthand{estonian}{~~}{\til}

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

```

Change History

estonian-1.0b	estonian-1.0c
General: corrected typos 1	\captionsestonian: Added \proofname for AMS-L ^A T _E X . . . 4

estonian-1.0d		<code>\~o</code> in T1 and incorrect display of <code>\~o</code> with the <code>times</code> package) .	6
General: The second argument was missing in the definition of some of the double-quote shorthands	7	removed definitions of macros <code>\@umlaut</code> and <code>\@tilde</code> as they seemed to have no purpose . . .	6
<code>\captionsestonian</code> : Added translation of ‘Proof’	4	removed macros <code>\dieresis</code> and <code>\texttilde</code> that were used to store the original definitions of <code>\" </code> and <code>\^</code> , as they would have no purpose now	6
<code>\noextrasestonian</code> : Removed the code that changes category, lower case, upper case and space factor codes	5	use <code>\allowhyphens</code> to allow hyphenation in words containing "a, "A, "o, "O, "u or "U	7
estonian-1.0e		use <code>\umlauthigh</code> to restore umlauts (before, <code>\babel@save\</code> was used but that did not work)	5
General: Now use <code>\ldf@finish</code> to wrap up	7	<code>\captionsestonian</code> : Added translation of ‘Glossary’	4
Now use <code>\LdfInit</code> to perform initial checks	4	<code>\et@gentilde</code> : do not redefine caron any more because the default one looks good enough . .	6
Replaced <code>\undefined</code> with <code>\@undefined</code> and <code>\empty</code> with <code>\@empty</code> for consistency with L ^A T _E X, moved the definition of <code>\atcatcode</code> right to the beginning.	1	renamed macros <code>\gentilde</code> and <code>\newtilde</code> to <code>\et@gentilde</code> and <code>\et@newtilde</code>	6
estonian-1.0f		use tilde for all letters except s and z (instead of using caron for all letters except o), like other <code>babel</code> language packages do (this fixes the display of ñ when using the <code>utf8</code> package) . .	6
General: Removed empty groups after double quote and guillemot characters	7	<code>\et@newtilde</code> : merged updates in the definition <code>\lower@umlaut</code> into <code>\et@newtilde</code> : removed <code>\allowhyphens</code> and added <code>\bgroup</code>	6
<code>\dateestonian</code> : use <code>\def</code> instead of <code>\edef</code>	4	estonian-1.1	
Use <code>\edef</code> to define <code>\today</code> to save memory	4	General: added usage guidelines to the documentation	1
estonian-1.0g		<code>\captionsestonian</code> : Replaced <code>\~o</code> with <code>\~o</code> for compatibility with the <code>notilde</code> option, and to have the same style, also replaced "u with <code>\"u</code> in “Peatükk”.	4
General: use <code>\bbl@t@one</code> instead of <code>\bbl@next</code>	6	<code>\noextrasestonian</code> : introduced the <code>notilde</code> option	5
estonian-1.0h		estonian-1.1a	
<code>\captionsestonian</code> : Added <code>\glossaryname</code>	4	General: small improvements in documentation	1
<code>\estonianhyphenmins</code> : Now use <code>\providehyphenmins</code> to provide a default value	5		
estonian-1.0j			
<code>\captionsestonian</code> : Replaced the translation of ‘Proof’	4		
estonian-1.0k			
General: corrected documentation of the commands <code>"-</code> and <code>\-</code> . . .	1		
redefine <code>\^</code> only for OT1 (before, redefining <code>\^</code> resulted in no hyphenation of words containing			