Babel support for Polish

current maintainer: Marcin Woliński

Version 1.3, 2022/03/12

1 Basic Usage

Typesetting Polish with \TeX\ is, surprisingly, more difficult than setting most other European languages. Polish uses several accented characters. Some of them, including ć, ń, ó, ś, ż, Ć, Ń, Ó, Ś, Ż, can be reasonably rendered by attaching “floating diacritics” to Latin letters. More problematic is the letter Ł, but legend has it that Donald Knuth needed to include the name \Lukasiewicz in his TAOCP, so he included a “diacritic” stroke in his Computer Modern fonts. Unfortunately, he had no reason to mention Wałęsa anywhere in TAOCP, and in result the “ogonek\textsuperscript{1}” is not present in original \TeX\ fonts. A different set of fonts is necessary to properly render Polish characters with ogonek: ą, ę, Ą, Ė (these characters are also used in Lithuanian along with ĭ, ĭ, Ĭ, Ĳ).

In Unicode \TeX\ engines (Xe\TeX, lua\TeX), Latin Modern fonts are used by default and these include a rich repertoire of accented Latin characters. With these engines, it’s enough to load Babel with the Polish module:

\usepackage[polish]{babel}

The matter is more complicated in the case of classical \LaTeX\ (including pdf\LaTeX). To obtain correct Polish letters, it is necessary to switch the document to a suitable font encoding. One possibility is:

\usepackage[OT4]{fontenc}
\usepackage[polish]{babel}

This will result in a document typeset using the Polish adaptation of Knuth’s classical Computer Modern fonts.

A more flexible option is the European font encoding T1, which covers most of European languages written in the Latin script. Unfortunately, the default fonts in this encoding are bitmap based, which will result in sub-optimal rendering. So with T1 it is necessary to select a different font family as well. A safe option is to

\textsuperscript{1}The Polish name seems to also be the official English name of the diacritic: https://en.wikipedia.org/wiki/Ogonek.
use Latin Modern (Latin Modern is practically identical to Computer Modern, a new name was needed to keep backward compatibility of Computer Modern):

\usepackage{lmodern}
\usepackage[T1]{fontenc}
\usepackage[polish]{babel}

Many font switching packages work with T1. Besides Latin Modern, you can expect well designed Polish characters to be present in all fonts from the \TeX{} Gyre collection (in particular the Times lookalike \texttt{tgtermes}, Bookmanish \texttt{tgbonum}, Century Schoolbook-like \texttt{tgschola}, etc.) and in other fonts originating from the Polish \TeX{} Users Group GUST\textsuperscript{2}. With Unicode engines you can also use any OpenType font present on your system (see the \texttt{fontspec} package).

It is also worth noting, that ogonek is not really an accent: it has to be separately designed for each base letter to get a typographically acceptable result. This is not the case, e.g., for the packages in the PSNFSS set, in particular \texttt{times.sty}. Although these support the T1 encoding, the ogonek protrudes from the characters at some random angle (this is particularly visible in the case of the letter ę):

\begin{tabular}{ll}
\texttt{\TeX{} Gyre Termes} & gąbkę gąbkę \\
\texttt{PSNFF Times} & gąbkę gąbkę
\end{tabular}

If quality is your goal, do not use \texttt{times.sty} for Polish, use \texttt{tgtermes.sty}.

Another element necessary for successful typesetting of Polish is a way to encode Polish accented characters in the source file. Fortunately, nowadays all \LaTeX{} engines assume input files to be in Unicode, in its UTF-8 form, so if you are on a modern system, this will work by default. (If you are forced to use some old byte-based encoding like the standard ISO 8859-2 or its Windows counterpart CP1250, please consult the documentation of the \texttt{inputenc} package.)

2 The Functions of this Module

As of version 1.3 the code has been updated to the new syntax provided by Babel to declare language specific elements. In its current form, the module is pretty typical \texttt{.ldf} file: it defines Polish translation of language specific strings used by \LaTeX{}, it ensures Polish hyphenation patterns are used, and activates the setting known as \texttt{\frenchspacing}.

Language modules of Babel typically check for the hyphenation patterns for the language in question. In their absence, a warning is issued and English hyphenation patterns are activated. We find this behaviour inappropriate. Polish hyphenation patterns are a part of standard \TeX{} distributions for over 20 years.

\textsuperscript{2}http://www.gust.org.pl/projects/e-foundry
Inability to load them signals some configuration problem in the system. Moreover, the use of English patterns would lead to very badly chosen breaks in a Polish text. Thus, we have decided to block automatic hyphenation in Polish text in this case. In consequence, system configuration has to be corrected or hyphenation points selected manually.

American typography has a rule, that space after punctuation characters is larger than typical inter-word space. This rule is by default followed by \TeX. In Polish tradition all spaces should be of equal width, so we activate \texttt{\frenchspacing} for Polish text.

Babel provides a mechanism for language-specific short commands (short-hands) consisting of the character " followed by one more character. Table 1 lists the commands available for Polish. The shorthand "- is similar to the command \-, but the former does not block automatic hyphenation in the rest of the word. The shorthand "=" should be used for explicit hyphens in the text. According to Polish rules, when a break occurs at an explicit hyphen, the hyphen gets repeated at the beginning of the new line. So, e.g., Polish compound adjectives like \texttt{biało-czerwony} 'white-and-red' should be input as \texttt{biało="czerwony}.

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
"- & soft hyphen \\
"= & explicit hyphen repeated after the break \\
\| & disable ligature at this position. \\
\" & Polish left double quote: " \\
\" & Polish right double quote: " \\
\"< & French left double quote: « (used in Polish as second level quotes) \\
\"> & French right double quote: » \\
\hline
\end{tabular}
\caption{Shorthands provided by the Polish module}
\end{table}

3 A Note on Compatibility

Incompatible changes were introduced in version 1.3 of this module. In case of problems with processing legacy documents, please try replacing the module invocation with the following:

\begin{verbatim}
\usepackage[polish-compat]{babel}
\end{verbatim}

This causes an exact copy of the module version v1.2l of 2005/03/31 to be loaded. However, please be aware that this old code has known problems (see Section 4.1). In compatibility mode a different list of shorthands is available, they are listed in Table 2.
Table 2: Shortcuts and commands provided in the compatibility mode. Be aware that these do not provide typographically acceptable shapes of Polish accented characters.

"a" or \aob, for a-ogonek (rendered a)
"A" or \Aob, for A-ogonek (rendered A)
"e" or \eob, for e-ogonek (rendered e)
"E" or \Eob, for E-ogonek (rendered E)
"c" or \c', for accented c (c), same with uppercase letters and n, o, s
"l" or \lpb{}, for l with stroke (l)
"L" or \Lpb{}, for L with stroke (L)
"r" or \zkb{}, for z-dot-accent (z), cf. pronunciation
"R" or \Zkb{}, for Z-dot-accent (Z)
"z" or \z', for accented z (z)
"Z" or \Z', for accented Z (Z)
"|" disable ligature at this position.
"-" an explicit hyphen sign, allowing hyphenation in the rest of the word.
"\" like ", but producing no hyphen sign (for compound words with hyphen, e.g. x-"y").
" for Polish left double quote (").
"" for Polish right double quote (").
"<" for French left double quote (\langle).
">" for French right double quote (\rangle).

4 The code

\LdfInit{polish}\captionspolish

The new code requires at least the version 3.63 of Babel where the macro \babel@texpdf was first introduced.
\@ifpackagelater{babel}{2021/08/30}{
\providecommand\babel@texpdf[4]{#1}%
}
\@ifpackagelater{babel}{2021/08/30}{%
\PackageError{polish.ldf}{Babel too old, at least version 2021/08/30 required}{%}
6 \providecommand\babel@texpdf[4]{#1}%
7 }

When this file is read as an option, i.e. by the \usepackage command, polish could be an ‘unknown’ language in which case we have to make it known. So we check for the existence of \l@polish to see whether we have to do something here.
\ifx\l@polish\undefined
9 \nopatterns{Polish}
10 \adddialect{\l@polish}{\unhyphenated}\fi
\captionspolish The macro \captionspolish defines all strings used in the four standard document classes provided with \LaTeX.
And now, the generic branch, using the LICR and assuming T1.
The macro \extraspolish will perform all the extra definitions needed for the Polish language. The macro \noextraspolish is used to cancel the actions of \extraspolish.

We specify that the polish group of shorthands should be used. These characters are ‘turned on’ once, later their definition may vary.

For Polish texts \frenchspacing should be in effect. We make sure this is the case and reset it if necessary.

4.1 Compatibility code

This section contains the code and documentation of version v1.2l (2005/03/31) of polish.ldf, which is used for polish-compat. We keep it as a compatibility fallback, since (more or less) this version was available for almost 20 years and some documents may depend on its idiosyncratic features. There are known problems with this code:
• The "-shorthands for Polish characters do not use LICR, so even with modern fonts the precomposed characters from the font are not used (in the case of q, ę, ł, ż).

• The substitute for ogonek accent defined for OT1 is not accessible with standard \k (besides being hideous).

• Wrong shape of ę is used in the word Część generated by the command \part.

• The code is evil: it redefines \selectfont, which leads to conflicts with some standard \LaTeX packages. Another known incompatibility is the macro \lll, which conflicts with AMS-\LaTeX.

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

\LdfInit{polish}\captionspolish

This file is part of the babel system, it provides the source code for the Polish language definition file. It was developed out of Polish.tex, which was written by Elmar Schalueck and Michael Janich. Polish.tex was based on code by Leszek Holenderski, Jerzy Ryll and J. S. Bień from Faculty of Mathematics, Informatics and Mechanics of Warsaw University, except of Jerzy Ryll (Instytut Informatyki Uniwersytetu Warszawskiego).

When this file is read as an option, i.e. by the \usepackage command, polish could be an ‘unknown’ language in which case we have to make it known. So we check for the existence of \l@polish to see whether we have to do something here.

\ifx\l@polish\undefined
\@nopatterns{Polish}
\adddialect\l@polish0\fi

The next step consists of defining commands to switch to (and from) the Polish language.

\captionspolish The macro \captionspolish defines all strings used in the four standard document classes provided with \LaTeX.

\addto\captionspolish{%
\def\prefacename{Przedmowa}%
\def\refname{Literatura}%
\def\abstractname{Streszczenie}%
\def\bibname{Bibliografia}%
\def\chaptername{Rozdzia\l}%
\def\appendixname{Dodatek}%
\def\contentsname{Spis tre\sci}%
\def\listfigurename{Spis rysunk\sci}%
\def\listtablename{Spis tablic}%
\def\indexname{Indeks}%
\def\figurename{Rysunek}%
}
The macro `\datepolish` redefines the command `\today` to produce Polish dates.

```
115 \def\datepolish{%
116 \def\today{%
117 \number\day˜\ifcase\month\or
118 stycznia\or lutego\or marca\or kwietnia\or maja\or
czerwca\or lipca\or
119 sierpnia\or wrzeńia\or pa\’dziernika\or listopada\or
grudnia\fi
120 \space\number\year}
121 }
```

The macro `\extraspolish` will perform all the extra definitions needed for the Polish language. The macro `\noextraspolish` is used to cancel the actions of `\extraspolish`.

For Polish 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 polish group of shorthands should be used.

```
121 \initiate@active@char{"}
122 \addto\extraspolish{\languageshorthands{polish}}
123 \addto\extraspolish{\bbl@activate{"}}
```

Don’t forget to turn the shorthands off again.

```
124 \addto\noextraspolish{\bbl@deactivate{"}}
```

The code above is necessary because we need an extra active character. This character is then used as indicated in table 2.

If you have problems at the end of a word with a linebreak, use the other version without hyphenation tricks. Some TeX wizard may produce a better solution with forcasting another token to decide whether the character after the double quote is the last in a word. Do it and let us know.

In Polish texts some letters get special diacritical marks. Leszek Holenderski designed the following code to position the diacritics correctly for every font in every size. These macros need a few extra dimension variables.

```
125 \newdimen\pl@left
126 \newdimen\pl@down
127 \newdimen\pl@right
128 \newdimen\pl@temp
```
\sob  The macro \sob is used to put the 'ogonek' in the right place.
\def\sob#1#2#3#4#5{\params: letter and fractions hl,ho,vl,vo
\setbox0\hbox{#1}\setbox1\hbox{$\mathchar'454$}\setbox2\hbox{p}\n\pl@right=#2\wd0 \advance\pl@right by-#3\wd1\n\pl@down=#5\ht1 \advance\pl@down by-#4\ht0\n\pl@left=\pl@right \advance\pl@left by\wd1\n\pl@temp=\pl@down \advance\pl@temp by\dp2 \dp1=\pl@temp\n\leavevmode\kern\pl@right\lower\pl@down\box1\kern-\pl@left #1}\n\aob  The ogonek is placed with the letters 'a', 'A', 'e', and 'E'.
\DeclareTextCommand{\aob}{OT1}{\sob a{.66}{.20}{0}{.90}}\n\DeclareTextCommand{\Aob}{OT1}{\sob A{.80}{.50}{0}{.90}}\n\DeclareTextCommand{\eob}{OT1}{\sob e{.50}{.35}{0}{.93}}\n\DeclareTextCommand{\Eob}{OT1}{\sob E{.60}{.35}{0}{.90}}\n\DeclareTextCommand{\eob}{T1}{\k e}\n\DeclareTextCommand{\Eob}{T1}{\k E}\n\DeclareTextCommand{\aob}{T1}{\k a}\n\DeclareTextCommand{\Aob}{T1}{\k A}\n\DeclareTextCommand{\eob}{T1}{\k e}\n\DeclareTextCommand{\Eob}{T1}{\k E}\n\DeclareTextCommand{\eob}{OT1}{\k e}\n\DeclareTextCommand{\Eob}{OT1}{\k E}\n
For the 'new' T1 encoding we can provide simpler definitions.
\DeclareTextCommand{\aob}{T1}{\UseTextSymbol{OT1}{\aob}}\n\DeclareTextCommand{\Aob}{T1}{\UseTextSymbol{OT1}{\Aob}}\n\DeclareTextCommand{\eob}{T1}{\UseTextSymbol{OT1}{\eob}}\n\DeclareTextCommand{\Eob}{T1}{\UseTextSymbol{OT1}{\Eob}}\n\DeclareTextCommand{\eob}{OT1}{\UseTextSymbol{OT1}{\eob}}\n\DeclareTextCommand{\Eob}{OT1}{\UseTextSymbol{OT1}{\Eob}}\n\\spb  The macro \spb is used to put the 'poprzeczka' in the right place.
\def\spb#1#2#3#4#5{\params: letter and fractions hl,ho,vl,vo
\setbox0\hbox{#1}\setbox1\hbox{\char'023}\n\pl@right=#2\wd0 \advance\pl@right by-#3\wd1\n\pl@down=#5\ht1 \advance\pl@down by-#4\ht0\n\pl@left=\pl@right \advance\pl@left by\wd1\n\ht1=\pl@down \dp1=-\pl@down\n\leavevmode\kern\pl@right\lower\pl@down\box1\kern-\pl@left #1}\n\skb  The macro \skb is used to put the 'kropka' in the right place.
\def\skb#1#2#3#4#5{\params: letter and fractions hl,ho,vl,vo
\setbox0\hbox{#1}\setbox1\hbox{\char'056}\n\pl@right=#2\wd0 \advance\pl@right by-#3\wd1\n\pl@down=#5\ht1 \advance\pl@down by-#4\ht0\n\pl@left=\pl@right \advance\pl@left by\wd1\n\leavevmode\kern\pl@right\lower\pl@down\box1\kern-\pl@left #1}\n\textpl  For the 'poprzeczka' and the 'kropka' in text fonts we don't need any special coding, but we can (almost) use what is already available.
Initially we assume that typesetting is done with text fonts.

But for the ‘teletype’ font in ‘OT1’ encoding we have to take some special actions, involving the macros defined above.

To activate these codes the font changing commands as they are defined in \texttt{LATEX} are modified. The same is done for plain \TeX{}’s font changing commands.

When \texttt{\selectfont} is undefined the current format is supposed to be either plain (based) or LATEX 2.09.

Currently we support the OT1 and T1 encodings. For T1 we don’t have to make a difference between typewriter fonts and other fonts, they all have the same glyphs.

For OT1 we need to check the current font family, stored in \texttt{\f@family}. Unfortunately we need a hack as \texttt{\ttdefault} is defined as a \texttt{\long} macro, while \texttt{\f@family} is not.
We save the original double quote character in \dq to keep it available, the math accent \" can now be typed as ".

\begin{verbatim}
begin\textcode{"}
def\textpl{end\textcode{"}}
def\dq{"}
\end{verbatim}

Now we can define the doublequote macros for diacritics,
\begin{verbatim}
\declare@shorthand{polish}{\"a}{\textormath{\aob}{\ddot a}}
\declare@shorthand{polish}{\"A}{\textormath{\Aob}{\ddot A}}
\declare@shorthand{polish}{\"c}{\textormath{\c}{\acute c}}
\declare@shorthand{polish}{\"C}{\textormath{\C}{\acute C}}
\declare@shorthand{polish}{\"e}{\textormath{\eob}{\ddot e}}
\declare@shorthand{polish}{\"E}{\textormath{\Eob}{\ddot E}}
\declare@shorthand{polish}{\"l}{\textormath{\lpb}{\ddot l}}
\declare@shorthand{polish}{\"L}{\textormath{\Lpb}{\ddot L}}
\declare@shorthand{polish}{\"n}{\textormath{\n}{\acute n}}
\declare@shorthand{polish}{\"N}{\textormath{\N}{\acute N}}
\declare@shorthand{polish}{\"o}{\textormath{\o}{\acute o}}
\declare@shorthand{polish}{\"O}{\textormath{\O}{\acute O}}
\declare@shorthand{polish}{\"s}{\textormath{\s}{\acute s}}
\ declare@shorthand{polish}{\"S}{\textormath{\S}{\acute S}}
\end{verbatim}

\texttt{\polishrz}

The command \polishrz defines the shorthands "r, "z and "x to produce pointed z, accented z and "x. This is the default as these shorthands were defined by this language definition file for quite some time.
\begin{verbatim}
\newcommand*{\polishrz}{%
\declare@shorthand{polish}{\"r}{\textormath{\zkb}{\ddot r}}%
\ declare@shorthand{polish}{\"R}{\textormath{\Zkb}{\ddot R}}%
\ declare@shorthand{polish}{\"z}{\textormath{\z}{\acute z}}%
\ declare@shorthand{polish}{\"Z}{\textormath{\Z}{\acute Z}}%
\ declare@shorthand{polish}{\"x}{\dq x}%
\ declare@shorthand{polish}{\"X}{\dq X}%
}\end{verbatim}

\texttt{\polishzx}

The command \polishzx switches to a different set of shorthands, "z, "x and "r to produce pointed z, accented z and "r; a different shorthand notation also in use.
\begin{verbatim}
\newcommand*{\polishzx}{%
\ declare@shorthand{polish}{\"z}{\textormath{\zkb}{\ddot z}}%
\ declare@shorthand{polish}{\"Z}{\textormath{\Zkb}{\ddot Z}}%
\ declare@shorthand{polish}{\"x}{\textormath{\z}{\acute x}}%
\ declare@shorthand{polish}{\"X}{\textormath{\Z}{\acute X}}%
\ declare@shorthand{polish}{\"r}{\dq r}%
\end{verbatim}
Then we define access to two forms of quotation marks, similar to the german and french quotation marks.

\declare@shorthand{polish}{"\textquoteleft}{\textquoteleft}
\declare@shorthand{polish}{"\textquotedblright}{\textquotedblright}
\declare@shorthand{polish}{"\guillemotleft}{\guillemotleft}
\declare@shorthand{polish}{"\guillemotright}{\guillemotright}

then we define two shorthands to be able to specify hyphenation breakpoints that behave a little different from \-.  
\declare@shorthand{polish}{"-}{\nobreak-\bbl@allowhyphens}
\declare@shorthand{polish}{"|}{\discretionary{-}{}{\kern.03em}}{}

And we want to have a shorthand for disabling a ligature.

\declare@shorthand{polish}{"|}{\nobreak-\bbl@allowhyphens}
\declare@shorthand{polish}{"|}{\discretionary{-}{}{\kern.03em}}{}

All that’s left to do now is to define a couple of commands for reasons of compatibility with polish.tex.

\def\mdqon{\shorthandon{"}}
\def\mdqoff{\shorthandoff{"}}

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.

\ldf@finish{polish}

("compat")