

# The Ukrainian Language in the `babel` system

Version 1.4d

Sergiy M. Ponomarenko\*

Released 2020/10/13

## Contents

<b>1</b>	<b>The Ukrainian Language Definition File</b>	<b>1</b>
<b>2</b>	<b>Usage</b>	<b>1</b>
2.1	<code>\ATeX</code> . . . . .	1
2.2	<code>Lua\ATeX</code> . . . . .	2
2.3	<code>Xe\ATeX</code> . . . . .	2
<b>3</b>	<b>User's commands</b>	<b>3</b>
3.1	Active character . . . . .	3
3.2	Math commands . . . . .	4
<b>4</b>	<b><code>T<sub>E</sub>X</code>nic details</b>	<b>4</b>
<b>5</b>	<b>Known problems</b>	<b>5</b>
<b>6</b>	<b>Implementation</b>	<b>5</b>
6.1	Initial setup . . . . .	5
6.2	Output encoding . . . . .	5
6.3	Input encoding . . . . .	8
6.4	Shorthands . . . . .	8
6.4.1	Quotes . . . . .	9
6.4.2	Emdash, endash and hyphenation sign . . . . .	9
6.5	Switching to and from Ukrainian . . . . .	10
6.5.1	Caption names . . . . .	10
6.5.2	Date in Ukrainian . . . . .	12
6.5.3	Hyphenation patterns . . . . .	13
6.5.4	Extra definitions . . . . .	13
6.6	Alphabetic enumerations . . . . .	13
6.7	Ukrainian mathematical typography traditions . . . . .	14
6.8	Final settings . . . . .	15

## 1 The Ukrainian Language Definition File

The file `ukraineb.ldf`<sup>1</sup> is the source file for the Ukrainian Language Definition file `ukraineb.ldf` to be loaded by the `babel` package with the option `ukrainian`. It's based on the Russian language definition file `russianb.ldf` derived by Igor A. Kotelnikov.

## 2 Usage

Typesetting Ukrainian texts implies that a special input and output encodings should be used. Input encodings are those which are used in source (`.tex`) file. Output encoding is also known as the font encoding. It is implemented within the font files.

---

\*E-mail: [sergiy\\_dot\\_ponomarenko\\_at\\_gmail\\_dot\\_com](mailto:sergiy_dot_ponomarenko_at_gmail_dot_com).

<sup>1</sup>The file described in this document has the version number 1.4d and was last revised on 2020/10/13.

Generally, the user may choose between different available Cyrillic encodings. The current support for Cyrillic uses LH family of MetaFont fonts and theirs Postscript versions such as **CM-super**. Lua $\text{\LaTeX}$  and Xe $\text{\LaTeX}$ , being the Unicode-based successors of  $\text{\LaTeX}$ , allow also for any Open Type (OTF) and True Type (TTF) fonts which has Cyrillic script, e.g. Computer Modern Unicode, Linux Libertine, and many other system fonts that came with Linux, Mac and Windows operating systems.

With the advent of Unicode,  $\text{\LaTeX}$  community are moving towards eliminating all existing encodings in favor of Unicode, but nowadays one should take care when switching from  $\text{\LaTeX}$  to Lua $\text{\LaTeX}$  or Xe $\text{\LaTeX}$  since different packages should be loaded for those compilers.

Since earlier versions **babel** did not support Xe $\text{\LaTeX}$  (at least for some languages including Ukrainian), the **polyglossia** package was generally recommended in the past for use with Xe $\text{\LaTeX}$  as a replacement for **babel**. Nowadays, **babel** can be used with any engines, including  $\text{\LaTeX}$ , PDF $\text{\LaTeX}$ , Lua $\text{\LaTeX}$ , and Xe $\text{\LaTeX}$ . Nevertheless some troubles may occur with some languages which have no promptly updated **.ldf** files.

## 2.1 $\text{\LaTeX}$

When user's document is compiled with **latex.exe** or **pdflatex.exe**, recommended set of packages includes the **inputenc** and **fontenc** packages. They should be loaded before **babel**, for example,

```
\usepackage[T1,T2A]{fontenc}
\usepackage[utf8]{inputenc}
\usepackage[english,ukrainian]{babel}
```

Some variations in the order of loading the packages are allowed in this case but it is better to follow one and the same convention at all circumstances: the **babel** package should go last, and **fontenc** must be the first.

Input encoding should be declared as option to the **inputenc** package. Known Cyrillic encodings include **cp866** (MS DOS), **cp1251** (Windows), **koi8-u** (UNIX) and their variants. Nowadays, this list is appended with **utf8** input encoding.

Output encodings (also known as font encodings) are declared as options to the **fontenc** package. Known Cyrillic encodings are **T2A**, **T2B**, **T2C**, **LCY**, and **X2**; **LWN** is excluded from Ukrainian support of **ukraineb.ldf** since **LWN** is excluded from the **cyrillic** bundle of related files.

## 2.2 Lua $\text{\LaTeX}$

If Unicode fonts are not available, Lua $\text{\LaTeX}$  can run in compatibility (8-bit) mode to use same font as  $\text{\LaTeX}$  does. However the package **inputenc** does not work with Lua $\text{\LaTeX}$  and should be substituted with **luainputenc**. Source file is to be converted to **UTF8** (Unicode-8) encoding; it is the only input encoding accepted by Lua $\text{\LaTeX}$ . The 8-bit mode is invoked by the following sequence of packages:

```
\usepackage[T1,T2A]{fontenc}
\usepackage[lut8]{luainputenc}
\usepackage[english,ukrainian]{babel}
```

The order of the packages is crucial for Lua $\text{\LaTeX}$  in 8-bit mode. Since both **luainputenc** and **babel** should know what is a selected font encoding, the **fontenc** package should be loaded first. Input encoding management for Lua $\text{\LaTeX}$  is needed only for compatibility with old documents. For new documents, using UTF-8 encoding and Unicode fonts is strongly recommended. *You've been warned!* See [tex.stackexchange.com/questions/31709/can-one-instruct-lualatex-to-use-t2a-encoded-fonts](https://tex.stackexchange.com/questions/31709/can-one-instruct-lualatex-to-use-t2a-encoded-fonts).

To invoke Unicode mode, one needs to load the **fontspec** package instead of **luainputenc** and **fontspec** and explicitly indicate which True Type or Open Type fonts should be used for roman, sans-serif and monospaced types. The following example shows how to load Computer Modern Unicode (CMU) fonts, which is a part of all modern  $\text{\LaTeX}$  distributions:

```
\usepackage{fontspec}
\defaultfontfeatures{Renderer=Basic,Ligatures={TeX}}
\setmainfont{CMU Serif}
\setsansfont{CMU Sans Serif}
\setmonofont{CMU Typewriter Text}
\usepackage[english,ukrainian]{babel}
```

The `\defaultfontfeatures` declares default font features for subsequent `\setmainfont` (which sets romanian fonts), `\setsansfont` (sans-serif) and `\setmonofont` (monospaced font). Font features can be set up on per font bases; for example

```
\usepackage{fontspec}
\setmainfont[Renderer=Basic,Ligatures={TeX}]{CMU Serif}
\setsansfont[Renderer=Basic,Ligatures={TeX,Historic}]{CMU Sans Serif}
\setmonofont{CMU Typewriter Text}
\usepackage[english,ukrainian]{babel}
```

Here `Renderer=Basic,Ligatures={TeX}` activates ligatures which are existed in  $\text{\LaTeX}$ .

Recall that the language enlisted last in the list of options of the `babel` package is assumed to be the main language of the document, which is also active language right after `\begin{document}`. As of version 3.9, the main language can be set as a value of the `main` option as follows

```
\usepackage{fontspec}
\usepackage[english,main=ukrainian,german]{babel}
```

## 2.3 Xe $\text{\LaTeX}$

In Xe $\text{\LaTeX}$ , there is also a special mode for 8-bit compatibility. One can use `\XeTeXinputencoding` to change the input encoding temporarily, and the `"bytes"` encoding makes Xe $\text{\LaTeX}$  works like a 8-bit  $\text{\LaTeX}$  engine:

```
\XeTeXinputencoding "bytes"
\usepackage[utf8]{inputenc}
\usepackage[T2A]{fontenc}
\usepackage[english,ukrainian]{babel}
```

Xe $\text{\TeX}$  can use a different input encoding but it always uses the Unicode internally, so that `\XeTeXinputencoding` performs a conversion of the input into Unicode; see [tex.stackexchange.com/questions/36188/do-xetex-and-luatex-always-use-unicode](https://tex.stackexchange.com/questions/36188/do-xetex-and-luatex-always-use-unicode).

Unicode mode is set up same way as for Lua $\text{\LaTeX}$ , however the option `Renderer=Basic` can be dropped:

```
\usepackage{fontspec}
\defaultfontfeatures{Ligatures={TeX}}
\setmainfont{CMU Serif}
\setsansfont{CMU Sans Serif}
\setmonofont{CMU Typewriter Text}
\usepackage[english,ukrainian]{babel}
```

## 3 User's commands

In a multilingual document, some typographic rules are language dependent and should apply to the whole document.

Regarding local typography, the macro `\selectlanguage{ukrainian}` switches to the Ukrainian language, with the following effects:

1. Ukrainian hyphenation patterns are made active;
2. `\today` prints the date in Ukrainian;
3. the caption names are translated into Ukrainian ( $\text{\LaTeX}$  only);
4. emdash typed by the ligature `---` might be 20% shorter when Ukrainian is the current language; the result depends on the current encoding; `---` always produce long emdash in Lua $\text{\TeX}$  and Xe $\text{\TeX}$  since these engines use same encodings for all languages;

5. emdash typed by the ligature ”--- in Ukrainian is 20% shorter, however the ligature ”--- might not be defined in other languages; a shorter emdash (i.e. `\cyrdash`) can be typeset in any language using special macros enlisted in table 1.

Additional commands are provided to typeset quotes:

1. French quotation marks can be entered using the commands `\guillemotleft` and `\guillemotright` which work in  $\text{\LaTeX} 2_\epsilon$  and  $\text{Plain}\text{\TeX}$ .
2. German quotation marks can be entered using the commands `\glqq` and `\grqq` which work in  $\text{\LaTeX} 2_\epsilon$  and  $\text{Plain}\text{\TeX}$ .

The macro `\Ukrainian` defined as an alias for `\selectlanguage{ukrainian}`, and its “opponent” `\English`, existed in `ukraineb.ldf` has been removed since the Ukrainian language definition file is wrong place for definition of macros which switch to a distinct other language.

The macro `\textcyrillic{<text>}` is intended to typeset small chunks of text in Ukrainian; it is essentially an alias for `\foreignlanguage{ukrainian}{<text>}`.

### 3.1 Active character

Table 1 shows macros and active string which can be used to typeset various dashes and quotes. In the Ukrainian language, the character ” is made active. It can be considered as second escape character in addition to `\`. Some dashes and all quotes can be typed using both active character ” and ordinary macros as indicated in the table. However, some shorthanded hyphenations have no macro counterpart.

Table 1: Extra definitions made by `ukraineb.ldf`

<code>\glqq</code>	“	German opening double quote (looks like „).
<code>\grqq</code>	”	German closing double quote (looks like “).
<code>\guillemotleft</code>	<<	French opening double quote (looks like <<).
<code>\guillemotright</code>	>>	French closing double quote (looks like >>).
<code>\dq</code>	”	Original quotes character (”).
<code>\babelhyphen{soft}</code>	”-	Optional (soft) hyphen sign, similar to <code>\-</code> but allows hyphenation in the rest of the word; equivalent to <code>\babelhyphen{soft}</code> in <code>babel 3.9</code> .
<code>\babelhyphen{empty}</code>	”	Similar to <code>”-</code> but prints no hyphen sign (used for compound words with hyphen, e.g. <code>x-”y</code> ); equivalent to <code>\babelhyphen{empty}</code> in <code>babel 3.9</code> .
<code>\babelhyphen*{nobreak}</code>	”~	Compound word mark without a breakpoint, prints hyphen prohibiting hyphenation at the point; equivalent to <code>\babelhyphen*{nobreak}</code> in <code>babel 3.9</code> .
<code>\babelhyphen{hard}</code>	”=	A compound word mark with a breakpoint, prints hyphen allowing hyphenation in the composing words. equivalent to <code>\babelhyphen{hard}</code> in <code>babel 3.9</code> .
<code>\babelhyphen{nobreak}</code>	”	Disables ligature at this position; equivalent to <code>\babelhyphen{nobreak}</code> (??) in <code>babel 3.9</code> .
<code>\cyrdash</code>		Row Cyrillic emdash (does not care spaces around).
<code>\cdash---</code>	”---	Cyrillic emdash in plain text.
<code>\cdash--~</code>	”--~	Cyrillic emdash in compound names (as in Mendeleev”--~Klapeiron).
<code>\cdash---*</code>	”--*	Cyrillic emdash for denoting direct speech.
	”,	Thin space (allows further hyphenation as in D.”,Mendeleev).

Note that the standard soft hyphen `\-` is equivalent to `\babelhyphen*{soft}`.

The quotation marks traditionally used in Ukrainian were borrowed from other languages (e.g., French and German) so they keep their original names.

The French quotes are also available as ligatures ‘<<’ and ‘>>’ in 8-bit Cyrillic font encodings (LCY, X2, T2\*) and in Unicode encodings (EU1 and EU2) and as ‘<’ and ‘>’ characters in 7-bit Cyrillic font encodings (OT2 and LWN).

In Unicode encodings EU1 and EU2 cyrdashes and quotes can be typed as single character if text editor makes it possible to insert characters which absent of standard keyboard. This method works as well for 8-bit fonts encoded according to T2A if source file is encoded with cp1251 or utf8.

By default, active double quote is switched on. It can be switched off any time using `\shorthandoff{’}` and the switched on again using `\shorthandon{’}`.

## 3.2 Math commands

The `ukraineb.ldf` defines few macros that can be used independently of current language. These are macros to be used in math mode to type the names of trigonometric functions common for Ukrainian documents: `\sh`, `\ch`, `\tg`, `\ctg`, `\arctg`, `\arcctg`, `\th`, `\cth`, and `\cosec`. Cyrillic letters in math mode can be typed with the aid of text commands such as `\textbf`, `\textsf`, `\textit`, `\texttt`, e.t.c.

## 4 T<sub>E</sub>Xnical details

The packages `inputenc` and `luainputenc` make Cyrillic letters active so that a compiler converts them into corresponding `\cyr<letter>` macro at compilation time. For example, Ukrainian letter ‘a’ matches macro `\cyra`, and capital Ukrainian letter ‘A’ matches `\CYRA`. The package `fontenc` then matches every macro `\cyr<letter>` to corresponding glyph in a font file depending on a declared font encoding.

Nowadays, Unicode makes `\cyr<letter>` macros outdated since both source file and font file are encoded consistently. These macros should therefore be removed because mixing them with Unicode characters breaks sorting mechanism of such utilities as `bibtex` and `makeindex`. For the sake of backward compatibility, `\cyr<letter>` are still kept for L<sup>A</sup>T<sub>E</sub>X, but they are bypassed if LuaL<sup>A</sup>T<sub>E</sub>X or XeL<sup>A</sup>T<sub>E</sub>X are detected.

## 5 Known problems

Before switching from a legacy 8-bit engine (`tex`, `pdftex`) to an Unicode engine (`xetex`, `luatex`) and vise versa delete all `.aux`, `.toc`, `.lot`, `.lof` files as they might have stored incompatible internal encodings.

## 6 Implementation

### 6.1 Initial setup

The macro `\LdfInit` performs a couple of standard checks that must be made at the beginning of a language definition file, such as checking the category code of the `@`-sign, preventing the `.ldf` file from being processed twice, etc.

```
1 \LdfInit{ukrainian}{captionsukrainian}
```

First, we check if LuaL<sup>A</sup>T<sub>E</sub>X or XeL<sup>A</sup>T<sub>E</sub>X is running. If so, we set boolean key `\if@ukr@uni@code` to true. It will be used to eliminate `\cyr...` commands, which were introduced in L<sup>A</sup>T<sub>E</sub>X2e to handle various Cyrillic input encoding. With the advent of Unicode L<sup>A</sup>T<sub>E</sub>X is moving to universal input encoding, so we consider these `\cyr...` commands as obsolete. They are preserved though for backward compatibility in case if L<sup>A</sup>T<sub>E</sub>X or PDFL<sup>A</sup>T<sub>E</sub>X are running.

We don’t load the `ifluatex` or `ifxetex` package because `\RequirePackage` is not allowed at the stage of processing options (note that `babel` loads this file right when it processes its own options) but we borrow code from these packages.

```
2 \ifdefined\if@ukr@uni@code
3   \PackageError{babel}{\if@ukr@uni@code already defined.\MessageBreak
4     Please contact author of ukraineb.ldf}
5   \relax
6 \fi
```

```

7 \newif\if@ukr@uni@code
8 \ifdefined\luatexversion
9   \@ukr@uni@odetrue
10 \else
11   \ifdefined\XeTeXrevision
12     \@ukr@uni@odetrue
13   \fi
14 \fi

```

Check if hyphenation patterns for the Ukrainian language have been loaded in `language.dat`. Namely, we check for the existence of `\l@ukrainian`. If it is not defined, we declare Ukrainian as dialect for the default language number 0 which almost for sure is English.

```

15 \ifx\l@ukrainian\@undefined
16   \@nopatterns{Ukrainian}
17   \adddialect\l@ukrainian0
18 \fi

```

Now `\l@ukrainian` is always defined.

## 6.2 Output encoding

We need to know font encoding that is supposed to be active at the end of the `babel` package. Default font encoding, set by  $\text{\LaTeX}$  core, is `OT1`. This can be changed by the `fontenc` package in case of  $\text{\LaTeX}$  and by `fontspec` package in case of  $\text{Lua}\text{\LaTeX}$ . It matters whether these packages are loaded before or after `babel`. In the latter case or if these packages are not loaded at all, `ukraineb.ldf` ignores their effect and tries to provide some reasonable settings. In particular, `T2A` will be selected for Ukrainian language if  $\text{\LaTeX}$  is running but `EU1` in case of  $\text{Xe}\text{\LaTeX}$  and `EU2` in case of  $\text{Lua}\text{\LaTeX}$ .

In Unicode mode, the package `fontspec` should be loaded instead of `fontenc` to make font preparation; `fontspec` loads the package `xunicode` which sets current encoding (kept in `\cf@encoding`) to `EU1` for  $\text{Xe}\text{\LaTeX}$  and `EU2` for  $\text{Lua}\text{\LaTeX}$ , and the `babel` package sets the macro `\latinencoding` to `\cf@encoding`. Since `babel` scan for value `\cf@encoding` within `\AtBeginDocument`, `\latinencoding` will be set to either `EU1` for  $\text{Xe}\text{\LaTeX}$  or `EU2` for  $\text{Lua}\text{\LaTeX}$  no matter which of the packages, `babel` or `fontspec` is loaded first.

### `\cyrillicencoding`

There is a limited list of encodings appropriate for Cyrillic text. We will look which of them is declared and keep its name in the macro `\cyrillicencoding`. Correct (but obsolete and now deleted) 7-bit Cyrillic encoding is `LWN`. Correct 8-bit Cyrillic encodings are `T2A` (default for 8-bit compilers), `T2B`, `T2C`, `LCY` and `X2`. Correct utf8 encodings are `TU` (default for  $\text{Xe}\text{\LaTeX}$  and  $\text{Lua}\text{\LaTeX}$ ), `EU1` (obsolete, formerly used for  $\text{Xe}\text{\LaTeX}$ ), `EU2` (obsolete, formerly used for  $\text{Lua}\text{\LaTeX}$ ).

In 8-bit ( $\text{\LaTeX}$ ) mode, user may choose between different non-unicode Cyrillic encodings—e.g., `X2` or `LCY`. If user wants to use another font encoding rather than default (`T2A`), he has to load the corresponding file *before* `babel.sty`.

Remember that for the Ukrainian language, the `T2A` encoding is better than `X2`, because `X2` does not contain Latin letters, and users should be very careful to switch the language every time they want to typeset a Latin word inside a Ukrainian phrase or vice versa.

We parse the `\cdp@list` containing encodings known to  $\text{\LaTeX}$  in the order 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`, `LCY`, `X2`, `T2C`, `T2B`, `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 ( $\text{Lua}\text{\LaTeX}$  and  $\text{Xe}\text{\LaTeX}$ ) since they can run in compatibility (8-bit) mode.

```

19 \def\@setcyrillicencoding{%
20   \def\sce@a##1##2{%
21     \edef\sce@b{##1}%
22     \edef\sce@c{##2}%
23     \ifx\sce@b\sce@c
24       \let\cyrillicencoding\sce@c
25     \fi}%
26   \def\cdp@elt##1##2##3##4{%
27     \sce@a{##1}{OT2}%
28     \sce@a{##1}{LCY}%
29     \sce@a{##1}{X2}%
30     \sce@a{##1}{T2C}%

```

```

31      \sca{##1}{T2B}%
32      \sca{##1}{T2A}%
33      \ifukr@uni@ode
34          \%sca{##1}{EU1}%
35          \%sca{##1}{EU2}%
36          \sca{##1}{TU}%
37      \fi}%
38      \cdp@list
39 }
40 \ifx\cyrillicencoding\undefined
41   \@setcyrillicencoding
42 \fi
43 \@onlypreamble\@setcyrillicencoding
44 \@onlypreamble\sca
45 \@onlypreamble\sca@b
46 \@onlypreamble\sca@c

```

The last lines are to free the memory occupied by the macros `\@setcyrillicencoding` and `\sca@x` that are useless in the document. The contents of `\@begindocumenthook` is cleared automatically.

If `\cyrillicencoding` is still undefined, we issue warning and provide reasonable default value for `\cyrillicencoding`. We then load default encoding definitions; we use the lowercase names (i.e., `lcyenc.def` instead of `LCYenc.def`) when we do that.

```

47 \ifdefined\cyrillicencoding
48 \else
49   \ifukr@uni@ode
50     \ifdefined\XeTeXrevision
51       \edef\cyrillicencoding{EU1}
52     \else
53       \ifdefined\luatexversion
54         \edef\cyrillicencoding{EU2}
55       \fi
56     \fi
57   \else
58     \edef\cyrillicencoding{T2A}
59   \fi
60   \PackageWarning{babel}%
61     {No Cyrillic font encoding has been loaded so far.\MessageBreak
62     A font encoding should be declared before babel.\MessageBreak
63     Default ‘\cyrillicencoding’ encoding will be loaded
64   }%
65   \lowercase\expandafter{\expandafter\input\cyrillicencoding enc.def\relax}%

```

As final wisdom, we repeat `\@setcyrillicencoding` at `\begin{document}` time. We could not avoid previous call to `\@setcyrillicencoding` since compiler scan `.aux` file before it executes delayed code, and `.aux` may contain `\set@language{ukrainian}`; the latter rises an error if `\cyrillicencoding` would not be defined by that time.

```

66   \AtBeginDocument{\@setcyrillicencoding}
67 \fi

```

`\Ukrainian`  
`\cyr`  
`\cyrillictext`

For the sake of backward compatibility we keep the macro `\Ukrainian` but redefine its meaning; now `\Ukrainian` is simply an alias for `\selectlanguage{ukrainian}`.

We define `\cyrillictext` and its alias `\cyr` but remove another alias `\Ukr`; these macros are intended for use within `babel` macros and do not perform complete switch of the language. In particular, they do no switch captions and the name of current language stored in the macro `\language`. This inconsistency might break some assumptions embedded into `babel`'s. For example, the `\iflanguage` macro will fail.

Second, `\cyrillictext` does not activate shorthands, so that `"<`, `">`, `"“`, `"”`, `"---`, e.t.c. will not work.

And third, `\cyrillictext` does not write its trace to `.aux` file, which might result in wrong typesetting of table of content, list of table and list of figures in multilingual documents.

Due to any of 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 `\Ukrainian` or the command `\foreignlanguage` defined in the `babel` core; their functionality is similar to `\selectlanguage{ukrainian}` but they did not switch caption names, dates and shorthands.



```

68 \DeclareRobustCommand{\Ukrainian}{\selectlanguage{\ukrainian}}
69 \DeclareRobustCommand{\cyrillictext}{%
70   \fontencoding\cyrillicencoding\selectfont
71   \let\encodingdefault\cyrillicencoding
72   \expandafter\set@hyphenmins\ukrainianhyphenmins
73   \language\l@ukrainian}%
74 \let\cyr\cyrillictext

```

NEXT PART OF CODE SHOULD BE MOVED TO `X2enc.def`, `X2enc.dfu`, IF NEEDED. Since the `X2` encoding does not contain Latin letters, we should make some redefinitions of  $\text{\LaTeX}$  macros which implicitly produce Latin letters.

Unfortunately, the commands `\AA` and `\aa` are not encoding dependent in  $\text{\LaTeX}$  (unlike e.g., `\oe` or `\DH`). They are defined as `\r{A}` and `\r{a}`. This leads to unpredictable results when the font encoding does not contain the Latin letters ‘A’ and ‘a’ (like `X2`).

```

75 \expandafter\ifx\csname T@X2\endcsname\relax\else
76   \DeclareTextSymbolDefault{\AA}{OT1}
77   \DeclareTextSymbolDefault{\aa}{OT1}
78   \DeclareTextCommand{\aa}{OT1}{\r a}
79   \DeclareTextCommand{\AA}{OT1}{\r A}
80 \fi

```

The macro `\cyrillictext` switches current (e.g., Latin) font encoding to a Cyrillic font encoding stored in `\cyrillicencoding`. The macro `\latintext` switches back. This method assumes that an font encoding is a Latin one. But in fact the latter assumption does not matter if any other language is switched on using same method, i.e. if corresponding `.ldf` file defines required macros to switch that language on from same standard (Latin) state. Since `\latintext` is defined by the core of `babel` we do not repeat its definition here.

```

81 %\DeclareRobustCommand{\latintext}{%
82 %   \fontencoding{\latinencoding}\selectfont
83 %   \def\encodingdefault{\latinencoding}}
84 %\let\lat\latintext

```

```
\textcyrillic {<text>}
```

The macros `\cyrillictext` and `\latintext` are declarations. For shorter chunks of text the commands `\textcyrillic` and `\textlatin` can be used.

The macro `\textcyrillic` takes an argument which is then typeset using the requested font encoding. It is thus an equivalent of `\foreignlanguage{\ukrainian}`.

```
85 \DeclareTextFontCommand{\textcyrillic}{\cyrillictext}
```

### 6.3 Input encoding

User should use the `inputenc` package when any 8-bit Cyrillic font encoding is used, selecting one of the Cyrillic input encodings. We do not assume any default input encoding, so the `inputenc` package should be explicitly called by `\usepackage{inputenc}` before `babel`. Note however that default font encoding T2A fits well enough to Ukrainian version of Windows ANSI encoding which is almost equivalent to cp1251.

```

86 \@ifpackageloaded{inputenc}{%
87   \if@ukr@uni@ode
88     \PackageWarning{babel}{inputenc should not be used with LuaTeX or XeTeX}
89   \fi
90 }{%
91   %\def\reserved@a{LWN}%
92   %\ifx\reserved@a\cyrillicencoding\else
93   \def\reserved@a{OT2}%
94     \ifx\reserved@a\cyrillicencoding
95     \else
96       \def\reserved@a{EU1}%
97       \ifx\reserved@a\cyrillicencoding\else
98         \def\reserved@a{EU2}%
99         \ifx\reserved@a\cyrillicencoding\else
100           \PackageWarning{babel}{
101             {No input encoding specified for Ukrainian language}
102           \fi

```



```

103         \fi
104     \fi
105     %\fi
106 }

```

## 6.4 Shorthands

The double quote character ” is declared to be active in Ukrainian language.

```
107 \initiate@active@char{”}
```

Initial activation state will set to on later in section 6.5.4.

**\dq** The active character ” is used as indicated in table 1. We save the original double quote character in the **\dq** macro to keep it available. The math accent ’ can now be typed as “”.

```

108 \begingroup \catcode‘\’12
109 \def\reserved@a{\endgroup
110     \def\@SS{\mathchar”7019 }
111     \def\dq{”}}
112 \reserved@a

```

### 6.4.1 Quotes

We set ‘ and ’ as shorthands for \quotedblbase and \textquotedblleft, respectively. These shorthands were defined through german quotes \glqq and \grqq, which in their turn are defined in babel.def via \quotedblbase and \textquotedblleft, respectively. It occurred, that old definition caused errors in Unicode mode if fontspec is loaded.

The shorthands ’< and ’> were declared to be equivalents for the French quotes \flqq and \frqq, respectively. They are defined in babel.def via \guillemotleft and \guillemotright. However, \flqq and \guillemotleft (and their right counterparts) are typeset differently if current encoding is not T1. Therefore, we define ’< and ’> directly through \guillemotleft and \guillemotright.

```

113 \declare@shorthand{ukrainian}{’<}{\quotedblbase}
114 \declare@shorthand{ukrainian}{’>}{\textquotedblleft}
115 \declare@shorthand{ukrainian}{’<}{\guillemotleft}
116 \declare@shorthand{ukrainian}{’>}{\guillemotright}

```

Next set of shorthands is intended for variations of standard macro \- which indicates explicitly breakpoint for hyphenation in a word. Meaning of these shorthands is explained in table 1.

Because of pdfstrings patches for ukrainian shorthands were removed from hyperref, the support for them was added to the ukrainian.ldf file.

```

117 \providecommand\texorpdfstring[2]{#1}
118 \declare@shorthand{ukrainian}{’~}{\hskip\z@skip}
119 \declare@shorthand{ukrainian}{’~}{%
120     \texorpdfstring{\textormath{\leavevmode\hbox{-}}{-}}{-}}
121 \declare@shorthand{ukrainian}{’=}{\nobreak-\hskip\z@skip}
122 \declare@shorthand{ukrainian}{’|}{%
123     \texorpdfstring{\textormath{\nobreak\discretionary{-}}{-}{\kern.03em}\allowhyphens}}{-}}

```

### 6.4.2 Emdash, endash and hyphenation sign

To distinguish between ”- and ”--- we must check whether the next after - token is a hyphen character. If it is, we output an emdash, otherwise a hyphen sign. Therefore T<sub>E</sub>X looks for the next token after the first ‘-’, writes its meaning to \ukrainian@sh@next and finally call for \ukrainian@sh@tmp.

```

124 \declare@shorthand{ukrainian}{’-}{%
125     \def\ukrainian@sh@tmp{%
126         \if\ukrainian@sh@next-\expandafter\ukrainian@sh@emdash
127         \else%
128             \expandafter\ukrainian@sh@hyphen%
129         \fi}%
130     \futurelet\ukrainian@sh@next\ukrainian@sh@tmp}

```

Two macros \ukrainian@sh@hyphen and \ukrainian@sh@emdash called by \ukrainian@sh@tmp are defined below. The second of them has two parameters since it must gobble next two hyphen signs.

```

131 \def\ukrainian@sh@hyphen{\nobreak-\bblallowhyphens}
132 \def\ukrainian@sh@emdash#1#2{\cdash-#1#2}

```

**\cdash** In its turn, \ukrainian@sh@emdash simply calls for \cdash which has rich use. It analyses 3rd of 3 characters and calls for one of few predefined macros \@Acdash, \@Bcdash, \@Ccdash.

```

133 \def\cdash#1#2#3{\def\tempx@{#3}%
134   \def\tempa@{-}\def\tempb@{~}\def\tempc@{*}%
135   \ifx\tempx@\tempa@\@Acdash%
136   \else%
137     \ifx\tempx@\tempb@\@Bcdash%
138     \else%
139       \ifx\tempx@\tempc@\@Ccdash%
140       \else%
141         \errmessage{Wrong usage of cdash}
142       \fi
143     \fi
144   \fi
145 }
```

All these 3 internal macros call for \cyrdash, which type Cyrillic emdash, but put different spaces around the dash.

\@Acdash is invoked by "---. It types Cyrillic emdash to be used inside a text and puts an unbreakable thin space before the dash if a *space* is placed before "--- in the source file; can be used after display maths formulae, formatted lists, enumerations, etc.

```

146 \def\@Acdash{\ifdim\lastskip>z@\unskip\nobreak\hskip.2em\fi
147   \cyrdash\hskip.2em\ignorespaces}%

```

\@Bcdash is invoked by "--~. It types Cyrillic emdash in compound names (like Mendeleev-Klapeiron); requires no space characters around and adds extra space after the dash.

```

148 \def\@Bcdash{\leavevmode\ifdim\lastskip>z@\unskip\fi
149   \nobreak\cyrdash\penalty\exhyphenpenalty\hskip\z@skip\ignorespaces}%

```

\@Ccdash is invoked by "--\*. It denotes direct speech and adds small space after the dash.

```

150 \def\@Ccdash{\leavevmode
151   \nobreak\cyrdash\nobreak\hskip.35em\ignorespaces}%

```

**\cyrdash** The \cyrdash macro is defined in Cyrillic font encodings (LCY, T2\*, OT2, and x2) by means of \DeclareTextSymbol. In T2\* encodings \cyrdash refers to same code point 22 as \textemdash does so that these two macros are equivalent. However the dash at the code point 22 have different length in different fonts. The dash in Cyrillic fonts LH is 20% shorter as compared to Latin fonts such as CM (Computer Modern). As a result, the dash typed by the ligature --- or its variations mentioned in Table ?? might change its length after \selectlanguage.

The \cyrdash macro is not available in Latin encodings such as T1. Therefor an explicit or implicit call for \cyrdash when current language is English causes an error. For such a case, we provide a fake default. A standard check such as \ifx\cyrdash\undefined ... \fi fails to detect absent definitions for Latin encodings since the \cyrdash macro is in fact defined. Therefor we use the \ProvideTextCommandDefault method:

```

152 \PackageInfo{babel}{Default for \string\cyrdash\space is provided}
153 %%\ProvideTextCommandDefault{\cyrdash}{\iflanguage{ukrainian}%
154 %%  {\hbox to.8em{--\hss--}}{\textemdash}}
155 \ProvideTextCommandDefault{\cyrdash}{\hbox to.8em{--\hss--}}

```

The \cyrdash macro is not defined in the Unicode encoding TU. The fake definition given above cope with this case.

Finally, we define a shorthand thin space to be placed between initials as in D.',Mendeleev. When used instead of \, as in D.\,Mendeleev it allows hyphenation in the next word.

```

156 \declare@shorthand{ukrainian}{',}{\nobreak\hskip.2em\ignorespaces}

```

## 6.5 Switching to and from Ukrainian

Now we define additional macros used to reset current language to Ukrainian and back to some original state. The package babel based on the assumption that original state is characterized by a Latin encoding. Previously, for back reset the macro \OriginalTeX was used, but now use \latintext for the same purpose.

### 6.5.1 Caption names

First, we define Ukrainian equivalents for Ukrainian caption names.

`\captionsukrainian` The macro `\captionsukrainian` defines caption names used in the four standard document classes provided with L<sup>A</sup>T<sub>E</sub>X. The macro `\cyr` activates Cyrillic encoding. It could be dropped if we would be sure that Ukrainian captions are called only if current language is Ukrainian. However, the macros such as `\Ukrainian` do not conform to strict rules of the package `babel` as explained in the above.

We now use `babel`'s 3.9 `\Set<macro>` macro for defining caption names as well as date. If Unicode engine is running, Cyrillic letters are typed in by their Unicode code-points.

```

157 %
158 % ----- Caption Names (Unicode case) -----
159 %
160 \if@ukr@uni@ode
161   \PackageInfo{ukrainian.ldf}{Executing the 3.9 or latter}
162   \StartBabelCommands*{ukrainian}{captions}[unicode, fontenc=EU1 EU2, charset=utf8]
163     \SetString\prefacename{Вступ}% [babel]
164     \SetString\refname{Література}% [only article]
165     \SetString\abstractname{Анотація}% [only article, report]
166     \SetString\bibname{Бібліографія}% [only book, report]
167     \SetString\chaptername{Розділ}% [only book, report]
168     \SetString\appendixname{Додаток}%
169     \SetString\contentsname{Зміст}%
170     \SetString\tocname{\contentsname}%
171     \SetString\listfigurename{Перелік ілюстрацій}%
172     \SetString\listtablename{Перелік таблиць}%
173     \SetString\indexname{Предметний покажчик}%
174     \SetString\authorname{Іменний покажчик}%
175     \SetString\figurename{Рис.}%
176     \SetString\tablename{Таблиця}%
177     \SetString\partname{Частина}%
178     \SetString\enclname{вкл.}%
179     \SetString\ccname{вих.}%
180     \SetString\headtoname{вх.}%
181     \SetString\pagename{с.}% [letter]
182     \SetString\seename{див.}%
183     \SetString\alsoname{див. \ також}%
184     \SetString\proofname{Доведення}% [amsthm]
185     \SetString\glossaryname{Словник термінів}%
186     \SetString\acronymname{Абревіатури}% [glossaries] {Acronyms}
187     \SetString\lstlistingname{Лістинг}% [listings] (the environment) {Listing}
188     \SetString\lstlistlistingname{Лістинги}% [listings] (the "List of") {Listings}
189     \SetString\nomname{Позначення}%
190     \SetString\notesname{Нотатки}% [endnotes] {Notes}

```

Additional definitions for the package `nomenc1`:

```

191 %
192 % ----- nomenc1 (Unicode case) -----
193 %
194   \ifdefined\nomname%
195     \addto\captionsukrainian{%
196       \def\eqdeclaration#1{, див.\nobreakspace(#1)}%
197       \def\pagedeclaration#1{, стор.\nobreakspace#1}%
198     }%
199   \fi
200   \EndBabelCommands
201 \else
202 %
203 % ----- Caption Names (Nonunicode case) -----
204 %
205   \StartBabelCommands*{ukrainian}{captions}
206     \SetString\prefacename{{\cyr\CYRV\cyr\cyr\cyr\cyr}}%
207     \SetString\refname{%
208       {\cyr\CYRL\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr}%
209     \SetString\abstractname{%
210       {\cyr\CYRA\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr}%
211     \SetString\bibname{%
212       {\cyr\CYRB\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr}%
213     \SetString\chaptername{{\cyr\CYRR\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr\cyr}}%

```

```

214 \SetString\appendixname{\cyr\CYRD\cyro\cyrd\cyra\cyrt\cyro\cyrk}}%
215 \SetString\contentsname{\cyr\CYRZ\cyrm\cyrii\cyrs\cyrt}}%
216 \SetString\tocname{\contentsname}%
217 \SetString\listfigurename{\cyr\CYRP\cyre\cyrr\cyre\cyrl\cyrii\cyrk
218 \ \cyrii\cyrl\cyryu\cyrs\cyrt\cyrr\cyra\cyrc\cyrii\cyrishrt}}%
219 \SetString\listtablename{\cyr\CYRP\cyre\cyrr\cyre\cyrl\cyrii\cyrk
220 \ \cyrt\cyra\cyrb\cyrl\cyri\cyrc\cyrsfts}}%
221 \SetString\indexname{\cyr\CYRP\cyrr\cyre\cyrd\cyrm\cyre\cyrt\cyrn\cyri\cyrishrt
222 \ \cyrp\cyro\cyrk\cyra\cyrz\cyrch\cyri\cyrk}}%
223 \SetString\authorname{\cyr\CYRII\cyrm\cyre\cyrn\cyrn\cyri\cyrishrt
224 \ \cyrp\cyro\cyrk\cyra\cyrz\cyrch\cyri\cyrk}}%
225 \SetString\figurename{\cyr\CYRR\cyri\cyrs.}}%
226 \SetString\tablename{\cyr\CYRT\cyra\cyrb\cyrl.}}%
227 \SetString\partname{\cyr\CYRCH\cyra\cyrs\cyrt\cyri\cyrn\cyra}}%
228 \SetString\enclname{\cyr\cyrv\cyrk\cyrl\cyra\cyrd\cyrk\cyra}}%
229 \SetString\ccname{\cyr\cyrk\cyro\cyrp\cyrii\cyrya}}%
230 \SetString\headtoname{\cyr\cyrv\cyrh.}}%
231 \SetString\pagename{\cyr\cyrs.}}%
232 \SetString\seename{\cyr\cyrd\cyri\cyrv.}}%
233 \SetString\alsoname{\cyr\cyrd\cyri\cyrv. \ \cyrt\cyra\cyrk\cyro\cyrz}}%
234 \SetString\proofname{\cyr\CYRD\cyro\cyrv\cyre\cyrd\cyre\cyrn\cyrn\cyrya}}%
235 \SetString\glossaryname{\cyr\CYRS\cyrl\cyro\cyrv\cyrn\cyri\cyrk
236 \ \cyrt\cyre\cyrr\cyrm\cyrii\cyrn\cyrii\cyrv}}%
237 \SetString\acronymname{\cyr\CYRA\cyrb\cyrr\cyre\cyrv\cyrii\cyra\cyrt\cyru\cyri\cyri}%
238 \SetString\lstlistingname{\cyr\CYRL\cyrii\cyrs\cyrt\cyri\cyrn\cyrg}%
239 \SetString\lstlistlistingname{\cyr\CYRL\cyrii\cyrs\cyrt\cyri\cyrn\cyrg\cyri}%
240 \SetString\nomname{\CYRP\cyro\cyrz\cyrn\cyra\cyrch\cyre\cyrn\cyrn\cyrya}%
241 \SetString\notesname{\CYRN\cyro\cyrt\cyra\cyrt\cyrk\cyri}%
242 \EndBabelCommands
243 \fi

```

## 6.5.2 Date in Ukrainian

`\dateukrainian` The macro `\dateukrainian` is used to reset the macro `\today` in Ukrainian.

```

244 %
245 % ----- Date (Unicode case) -----
246 %
247 \if@ukr@uni@ode
248 \PackageInfo{ukrainian.ldf}{Executing the post 3.9 branch for dates}
249 \StartBabelCommands*{ukrainian}{date}[unicode, fontenc=EU1 EU2, charset=utf8]
250 \SetStringLoop{month#1name}{%
251     січня,%
252     лютого,%
253     березня,%
254     квітня,%
255     травня,%
256     червня,%
257     липня,%
258     серпня,%
259     вересня,%
260     жовтня,%
261     листопада,%
262     грудня%
263 }
264 \SetString\abbgyear{p.}%
265 \else
266 %
267 % ----- Date (Nonunicode case) -----
268 %
269 \StartBabelCommands*{ukrainian}{date}
270 \SetStringLoop{month#1name}{%
271     \cyrs\cyrii\cyrch\cyrn\cyrya,%
272     \cyrl\cyryu\cyrt\cyro\cyrg\cyro,%
273     \cyrb\cyre\cyrr\cyre\cyrz\cyrn\cyrya,%
274     \cyrk\cyrv\cyrii\cyrt\cyrn\cyrya,%

```



Nevertheless, the Ukrainian `babel` by default turns on alphabetical enumeration with Cyrillic letters. This means that enumerated lists that would be labelled with Latin letters in Latin scripts are labelled with Cyrillic ones instead.

**\Alph** Starting from this version, we remove an macro `\Asbuk` (also its lowercase counterpart `\asbuk`). Now we redefine the macro `\Alph`, which now produces (uppercase) Cyrillic letters instead of Latin ones when Ukrainian is switched on.

The letters `Ґ`, `Є`, `З`, `І`, `Ї`, `Й`, `О`, `Ч`, `Ь`, are skipped for such enumeration (see [ДСТУ 3008:2015](#)).

```
301 \addto\extrasukrainian{%
```

When Ukrainian switched off, the previous meaning of `\@Alph` will be restored

```
302 \babel@save{\@Alph}%
303 \if@ukr@uni@ode%
304 \def\@Alph#1{%
305 \ifcase#1\or%
306 А\or Б\or В\or Г\or Д\or Е\or Ж\or%
307 И\or К\or Л\or М\or Н\or П\or Р\or%
308 С\or Т\or У\or Ф\or Х\or Ц\or Ш\or%
309 Щ\or Ю\or Я%
310 \else%
311 \@ctrerr%
312 \fi}%
313 \else
314 \def\@Alph#1{%
315 \ifcase#1\or%
316 \CYRA\or\CYRB\or\CYRV\or\CYRG\or\CYRD\or\CYRE\or\CYRZH\or%
317 \CYRI\or\CYRK\or\CYRL\or\CYRM\or\CYRN\or\CYRP\or\CYRR\or%
318 \CYRS\or\CYRT\or\CYRU\or\CYRF\or\CYRH\or\CYRC\or\CYRSH\or%
319 \CYRSHCH\or\CYRYU\or\CYRYA%
320 \else%
321 \@ctrerr%
322 \fi}%
323 \fi
324 }
```

**\alph** Now the macro `\alph` produces lowercase Cyrillic letters.

The lowercase letters `ґ`, `є`, `з`, `і`, `ї`, `й`, `о`, `ч`, `ь`, are also skipped such enumeration (see [ДСТУ 3008:2015](#)).

```
325 \addto\extrasukrainian{%
```

When Ukrainian switched off, the previous meaning of `\@alph` will be restored

```
326 \babel@save{\@alph}%
327 \if@ukr@uni@ode%
328 \def\@alph#1{%
329 \ifcase#1\or%
330 а\or б\or в\or г\or д\or е\or ж\or%
331 и\or к\or л\or м\or н\or п\or р\or%
332 с\or т\or у\or ф\or х\or ц\or ш\or%
333 щ\or ю\or я%
334 \else%
335 \@ctrerr%
336 \fi}%
337 \else
338 \def\@alph#1{%
339 \ifcase#1\or%
340 \cyra\or\cyrb\or\cyrv\or\cyrg\or\cyrd\or\cyre\or\cyrzh\or%
341 \cyri\or\cyrk\or\cyrll\or\cyrml\or\cyrn\or\cyrp\or\cyrre\or%
342 \cyrs\or\cyrtr\or\cyrul\or\cyrfl\or\cyrh\or\cyrcl\or\cyrsh\or%
343 \cyrshch\or\cyrul\or\cyrul\or\cyrul%
344 \else%
345 \@ctrerr%
346 \fi}%
347 \fi
348 }
```

## 6.7 Ukrainian mathetematical typography traditions

`\sh` We also define few math operator names according to Ukrainian mathetematical typography traditions. Some math functions in Ukrainian math books have names different from English writings. For example, `\sinh` in Ukrainian is called `sh`. Special consideration needs the macro `\th` that conflicts with the text symbol `\th` defined in Latin 1 encoding:

```
\arctg 349 \def\sh{\mathop{\operator@font sh}\nolimits}
\arcctg 350 \def\ch{\mathop{\operator@font ch}\nolimits}
\th      351 \def\tg{\mathop{\operator@font tg}\nolimits}
\ctg     352 \def\ctg{\mathop{\operator@font ctg}\nolimits}
\cosec   353 \def\arctg{\mathop{\operator@font arctg}\nolimits}
         354 \def\arcctg{\mathop{\operator@font arcctg}\nolimits}
         355 \addto\extrasukrainian{%
         356     \babel@save{\th}%
         357     \let\ltx@th\th
         358     \def\th{\textormath{\ltx@th}%
         359         {\mathop{\operator@font th}\nolimits}}}%
         360 }
         361 \def\cth{\mathop{\operator@font cth}\nolimits}
         362 \def\cosec{\mathop{\operator@font cosec}\nolimits}
```

For compatibility with older Ukrainian packages we leave definition of the `\No` macro. However the Ukrainian number sign is now superseded with `\textnumero`. Moreover, it can be found on the keyboard.

```
363 \DeclareRobustCommand{\No}{%
364     \ifmmode{\nfss@text{\textnumero}}\else\textnumero\fi}
```

## 6.8 Final settings

The macro `\ldf@finish` does work needed at the end of each `.ldf` file. This includes resetting the category code of the `@`-sign, loading a local configuration file, and preparing the language to be activated at `\begin{document}` time.

```
365 \ldf@finish{ukrainian}
```