

Notes on Bulgarian typesetting in \LaTeX

Georgi Boshnakov

21 май 2006 г.

1 Introduction

This document shows some features specific to the Bulgarian typesetting with `babel`. The text is mainly English but the main language of the document, as far as \LaTeX is concerned, is Bulgarian. This is possible because the encoding T2A contains the latin characters. On the other hand, the effect is that English hyphenation is not turned on and therefore English words are not hyphenated or, if they are, they may be wrongly hyphenated. I have done this deliberately to emphasise that it is a good practice to change explicitly the language.

This document was written initially for my package `bulgaria` developed in 1994-1996 to support Bulgarian language in the then new version of \LaTeX . Since at the time the encoding systems in general were in transition and because there was no standard for cyrillic ones I assembled some fonts to support the MIK encoding (the prevalent encoding at the time in Bulgaria, its origins are in DOS but it was supported in Windows as well, not by Microsoft though). This allowed me to introduce a fully working implementation only a few months after \LaTeX 2e became the official \LaTeX version.

I later modified this to work with the Babel system the most important change being the switch to the use of standard fonts and encodings. I handed this to the Babel developer Johannes Braams who is currently maintaining it.

Note that currently the MikTeX suite replaces my hyphenation patterns (available as part of the `babel` package) with different ones which seem not to work. So, I release my hyphenation files separately as well. These can be found in the `language` subdirectory on CTAN or on my web page.

Any comments and suggestions are welcome. I do not give my email address here (because of the spamming curse) but you can easily find me with a search engine. I will maintain the hyphenation and, if there is demand,

release a new set of patterns to support the more stylish traditional hyphenation (the current patterns are the simplified ones, born in the computer age).

2 Language and encoding

There are still many encoding systems around, you need to tell \LaTeX which one you use. For example,

```
\usepackage[cp1251]{inputenc}
```

specifies the cyrillic encoding typically used in Windows. This line should be one of the first few lines after the `\documentclass` declaration.

Note that the input encoding is completely independent of the internal encoding used by \LaTeX . You generally do not need to be concerned with it but for completeness here is the command that specifies the internal (cyrillic) encoding T2A:

```
\usepackage[T2A]{fontenc}
```

This should not be needed but see the section about uppercase/lowercase below.

To change language use may use the babel assortment of commands, say

```
\selectlanguage{english}  
\selectlanguage{bulgarian}
```

see the user guide of Babel for details.

There are useful shorthands like `\Bul`, `\Bg`, `\Eng`, `\latintext`, `\cyrillictext`.

3 Alphabet

3.1 Entering cyrillic letters

The best way is to use some of the standard encodings and leave \LaTeX to struggle with the conversion work (with knows most of the cyrillic encodings). Put a line like

```
\usepackage[MIK]{inputenc}
```

near the beginning of your document to tell the input encoding (MIK in this example). This command is usually placed write after the `\documentclass` command.

If you need to enter only a couple of words you may also use commands like `\cyrA`, `CYRA`, `\cyryu`, `\CYRYU`. The number sign can be obtained by `\No` or `\textnumero`.

For example, `\CYRA\CYRB\CYRV`, `\cyrA\cyrB\cyrV` produces

АБВ, абв

Accents in Bulgarian are used very rarely. This is done only to designate where the stress falls in certain special circumstances (mainly to avoid confusion). For example, `\‘и` gives `й`. This is unambiguously a pronoun which if written without the accent may occasionally be confused with the conjunction `и` (Bulgarian for “and”).

3.2 Changing letter case (Uppercase/Lowercase)

Use the \LaTeX commands `\MakeLowercase` and `\MakeUppercase` to change case of a piece of text. For example,

`\MakeLowercase{АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ №}`

gives

абвгдежзийклмнопрстуфхцчшщъыьэюя №

(All letters should change to lowercase, if `\MakeLowercase` works properly.)

Similarly

`\MakeUppercase{абвгдежзийклмнопрстуфхцчшщъыьэюя №}`

gives

АБВГДЕЖЗИЙКЛМНОПРСТУФХЦЧШЩЪЫЬЭЮЯ №

(All letters should change to uppercase, if `\MakeUppercase` works properly.)

Note that the symbol № is unchanged by both commands since it is not a regular letter.

If the above commands do not change the case properly, then this should be rectified by including the command `\usepackage[T2A]{fontenc}` in the preamble of your document (this command should not be needed as T2A is the default encoding, a bug may have crawled in somehow). If you have the `.tex` source of this file you may wish to comment out the `\usepackage` command for package `fontenc` and compile the file to see the difference, if any.

3.3 The symbol №

The symbol № is not particularly important but it is used occasionally, e.g.,

№ 1, № 20, №№ 1–5.

The Bulgarian keyboard has a key №. The commands `\textnumero` and `\No` produce № as well.

LaTeX source	typeset output
<code>\textnumero\,3</code>	№ 3
<code>\No\,3</code>	№ 3
<code>\# \,3</code>	№ 3

3.4 Roman numbers

In Bulgarian texts, uppercase Roman numbers uppercase are often used for ordinal numbers. Most of them can be entered via the Bulgarian keyboard: the letters I and V are available precisely for that purpose, M and C are the same in the cyrillic alphabet, but L is missing.

The command `\Romannumeral` (with shorthand `\#` defined in the preamble of this file) is provided for capitalised Roman numbers, e.g.: `\#1`, `\#8`, `\#26`, `\#{26}`, give I, VIII, II6, and XXVI, respectively. This example shows also that arguments with 2 or more decimal digits have to be put in braces.

3.5 Quoting styles

The quotation marks in texts typeset in Bulgarian traditionally look like „this“. This style and the quotes themselves have been borrowed from the German language.

To get traditional Bulgarian style quotes enclose the text in "‘ and "’". To get the French style quotes use "< and "> instead. For example,

```
‘‘English style quotes’’ \par
“traditional Bulgarian style quotes” \par
“<Alternative Bulgarian style quotes in older texts”>
```

gives

```
“English style quotes”
„traditional Bulgarian style quotes“
«alternative Bulgarian style quotes in older texts»
```

Note that in the pre-babel `bulgarian.sty` quotes in Bulgarian texts were typeset automatically in the traditional Bulgarian style.

3.6 Dates

The command `\today` produces the current date, e.g. today is 21 май 2006 г..

In the traditional format for dates in Bulgaria uppercased roman numbers are used for the months, e.g. today is 21. V. 2006 г..

\LaTeX source	typeset output
<code>\today</code>	21 май 2006 г.
<code>\todayRoman</code>	21. V. 2006 г.

4 Hyphenation

\LaTeX uses separate hyphenation rules for each language. This happens behind the scene when you change the language with the designated commands. This is done using hyphenation patterns incorporated in a \LaTeX “format” created during the installation of the \TeX system.

If your Bulgarian text does not get hyphenated then you will need to re-create the “formats” with Bulgarian support. This is fairly simple, see the documentation of your \TeX distribution for details.

If this does not work, fetch the hyphenation patterns for Bulgarian from CTAN or my web page and follow the instructions therein. You will probably have to do this if you are using MikTeX, the major Windows implementation of \TeX , since currently the Bulgarian support seems to be broken in the MikTeX distribution. (Requests for remedy prompted me to revise this and some other files.)

5 Enumeration

Alphabetic enumeration is usually done with the Bulgarian alphabet but latin enumeration is used as well. Besides the alphabet, the style of the enumeration is somewhat different from the English one.

We provide facilities for both cyrillic and latin enumeration. By default, the cyrillic one is switched on by commands changing the language to Bulgarian. The declarations `\alphEng` and `\alphBul` may be used to switch to and from latin enumeration, if desired.

Note that the cyrillic enumeration here simply uses cyrillic letters wherever the standard \LaTeX enumeration style would produce alphabetic enumeration. A fuller implementation might change other details but this is better left to document classes.

The letters `ishrt`, `snhrd` and `sftsn`, are not used for enumeration and therefore omitted. Note that the letters `erev` and `eri` are not part of the Bulgarian alphabet anyway.

1. a
 - (a) aa

- i. aaa
 - A. aaaa
 - Б. bbbb
 - B. cccc
 - Г. dddd
 - ii. bbb
- (б) bb
- (в) cc
- (г) dd

2. b

The following enumeration is preceded by the command `\enumEng`.

1. a
- (a) aaaaaaa
 - i. aaaaaaa
 - A. aaaaaaa
 - B. bbbbbbb
 - C. cccccc
 - ii. bbbbbbb
 - iii. cccccc

2. b