

# Notes on Bulgarian typesetting in $\text{\LaTeX}$

Georgi Boshnakov

1 юли 2008 г.

## Съдържание

1	Introduction	1
2	Language and encoding	2
3	Alphabet	3
3.1	Entering cyrillic letters . . . . .	3
3.2	Changing letter case (Uppercase/Lowercase) . . . . .	4
3.3	The symbol № . . . . .	4
3.4	Roman numbers . . . . .	4
3.5	Quoting styles . . . . .	5
3.6	Dates . . . . .	5
4	Hyphenation	6
5	Enumeration	6
6	Support for the varioref package	7
7	Reporting bugs and requesting features	8

## 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  $\text{\LaTeX}$  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 `LATEX2e` 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.

Prompted by some enquiries, I released the hyphenation stuff separately in may 2006. It can be found in the `language` subdirectory on CTAN or on my web page.

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.

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 you 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`, and `\cyrillictext`.

## 3 Alphabet

### 3.1 Entering cyrillic letters

The best way to enter Bulgarian text is to type it in some of the standard encodings and leave  $\LaTeX$  to struggle with the conversion work (it knows most of the cyrillic encodings). To tell  $\LaTeX$  the input encoding, put near the beginning of your document a line like

```
\usepackage[utf8]{inputenc}
```

if you are in the modern age and/or Linux, or

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

if you prefer “classic DOS” and the MIK encoding, or

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

if you do not know what I am talking about and are on a Windows system.

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`, `\cyrju`, `\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 `и́`. This is unambiguously a pronoun which if written without the accent may occasionally be confused with the conjunction `и` (Bulgarian for “and”).

Warning: the shorter `\‘и` works in all encodings except `utf8`. It is prudent to always add the braces to avoid irritation if you decide to switch to `utf8`.

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

### 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 №
```

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.

### 3.4 Roman numbers

In Bulgarian texts, uppercase Roman numbers 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. This does not matter much in the case of a computer keyboard. Even so, I prefer to redefine the I and V keys for more

important tasks and leave the computer render a number in Roman form, see also the comments in section `sec:bugs`.

The command `\Romannumeral` (with shorthand `\R#` defined in the preamble of this file) is provided for capitalised Roman numbers, e.g.: `\R#1`, `\R#8`, `\R#26`, `\R#{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
"‘текст в традиционните за български текстове кавички”’ \par
"<алтернативни кавички, срещани предимно в по-стари текстове">
```

gives

```
“English style quotes”
„текст в традиционните за български текстове кавички“
«алтернативни кавички, срещани предимно в по-стари текстове»
```

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 1 юли 2008 г..

In the traditional format for dates in Bulgaria uppercased roman numbers are used for the months, e.g. today is 1. VII. 2008 г..

$\LaTeX$ source	typeset output
<code>\today</code>	1 юли 2008 г.
<code>\todayRoman</code>	1. VII. 2008 г.

## 4 Hyphenation

When you choose a language for `babel`,  $\LaTeX$  hyphenates the text according to the hyphenation rules of that language. So, normally you do not need to do anything more to get proper hyphenation.

If your Bulgarian text does not get hyphenated, then you need to tell the  $\TeX$  system to load Bulgarian hyphenation rules. The details are system dependent but not complicated, see the documentation of your  $\TeX$  distribution for details.

For example, if you are using MikTeX, start its Settings utility (known also as MikTeX options, click the Language tab, tick `bulgarian`, and click OK. MikTeX will tell you that it needs to regenerate the formats, let it do it. That is it. Process a document to see that hyphenation works.

If you find words that do not hyphenate properly, then send me an email.

## 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. This is definitely not satisfactory and a fuller implementation might change other details but this is better left to document classes.

The letters `й`, `ъ`, and `ь` are not used for enumeration and therefore omitted. Note that the cyrillic letters `э` and `ы` 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. ccccc
    - ii. bbbbbbb
    - iii. ccccc
2. b

## 6 Support for the varioref package

The package `varioref` allows for producing cross-references which look natural and generally do not require manual intervention when the context changes. Support for Bulgarian should be available with the next release of `LATEX`. This file has been produced with a draft version of the package.

The table below is an illustration. Notice that the page reference command is the same in all cases but the printed text changes depending on its relative distance from the reference point. Also, references to the current page and the facing page are automatically varied to some extent.

command	typeset output
<code>\vpageref{sec:intro}</code>	на страница 1
<code>\vpageref{enum:1}</code>	на текущата страница
<code>\vpageref{enum:2}</code>	на тази страница
<code>\vpageref{table:No}</code>	на страница 4

It is instructive to see what output would have produced if the above table turned up on the following page. The table on the next page is generated by the same `LATEX` source as the table above. There is a `\newpage` command after this paragraph to ensure that the two tables are on different pages.

command	typeset output
<code>\vpageref{sec:intro}</code>	на страница 1
<code>\vpageref{enum:1}</code>	на предишната страница
<code>\vpageref{enum:2}</code>	на предната страница
<code>\vpageref{table:No}</code>	на страница 4

## 7 Reporting bugs and requesting features

Please email me bug reports and requests for features, visit my web page at the University of Manchester for further information.

I am using Emacs for typesetting documents. Emacs 22 has good support for essentially all writing systems in the world. AucTeX is an amazing Emacs package which makes an excellent environment for writing LaTeX (and other “dialects” of TeX) documents). Another Emacs package, ESS (Emacs speaks statistics), integrates Emacs smoothly with some of the major statistical systems.

If you are using Emacs but have not considered AucTeX yet, do it! If you do computations with systems such as R, S-plus or SAS, consider ESS.

If you are not an Emacs user you may consider becoming one. Together with the above tools it may become your basic working environment. Be prepared though for a period of familiarisation with the idiosyncracies of Emacs, it will help if you have a friend with some experience with it.