

A Russian style for Babel: problems and solutions

Olga Lapko and Irina Makhovaya

Mir Publishers
2, Pervyi Rizhskii Pereulok
Moscow, 129820
Russia
irina@mir.msk.su

Abstract

As with other languages using nonlatin basis there are some typographic features and national peculiarities that must be shown in the style. The paper describes the Russian style with macros `\captionrussian` for four standard Russian documents, `\daterussian`, `\Asbuk` and `\asbuk` for Russian alphabet counters and `\mathrussian` for Russian math operators. Some problems concerning the usage of this style (e.g. usage of different encodings) are described.

1 Introduction

As is generally known, \TeX is based on Latin alphabet and theoretically it is possible to use \TeX for other alphabets: Greek, Arabic, Cyrillic and so on. But there are a lot of troubles when we try to use \TeX for other alphabets in practice. Babel package is the first successful attempt to solve the problems of multilingual \TeX .

In this paper we discuss the concrete difficulties we encountered when creating the `russianb`¹ file for Babel. There are a lot of typographic features in Russian documents that can be separated into 3 classes:

1. the features that were borrowed from European typography, especially German and French;
2. the features that are peculiar for Russian typography only and there are no problems to describe them in the file `russianb`;

1. The filename of Babel style for Russian language is `russianb` as an analog of `germanb` to avoid probable confusion with \LaTeX 2.09 versions. Now the `russianb` is a beta-version – part of *CyrTUG-emTeX* appendix.

3. the features that are peculiar for Russian typography and there are some difficulties in describing and using them.

We also see two very important problems: variety of encoding schemes and necessity of portability of this file to different platforms, which we can solve only partially.

Now we shall describe the macros of the file `russianb` according to the classification.

2 Macros that were borrowed from other styles

The file `russianb` was derived initially from the original versions of `german` (for $\text{\LaTeX}2.09$) and `germanb` (for $\text{\LaTeX}2_{\epsilon}$) and `francais` (for $\text{\LaTeX}2_{\epsilon}$). These files have the language-specific macros which Russian typographic rules need:

1. *from* `germanb`
 - macros for French and German double quotes. *Note:* French double quotes are created by METAFONT in Cyrillic font and have their own ligature (e.g. <<)
 - “shorthands” for hyphenation in compound words and words with nonliteral characters (Russian words are not so long as German ones but rather long too). As in `germanb` the sign " was done active, certainly.
 - `\lefthyphenmin`–`\righthyphenmin`: for the Russian (as well as for the German) language hyphenation patterns are used with values 2–2;
2. *from* `francais`
 - macros for `:`, `,`, `?`, `!` signs: the amount of white space is increased before these signs: \TeX looks for a space between word and this sign, then, if it is, \TeX “unskips” it and places a little white space about 0.1em. *Note:* in `francais` \TeX places such space in case of space between word and sign and the amount of this space is somewhat larger;
 - `\frenchspacing` is switched on;
 - some additional signs (as well as in `francais`) are described in our style, e.g. number sign.

3 Macros that are created in Russian style and have no problems in usage

There are some macros borrowed from `russian.sty` (for $\text{\LaTeX}2.09$) of different releases:

1. macros for math operators whose names differ from English ones (e.g. in Russian manuscripts we write `tg` and `ctg` instead of `tan` and `cotan`);
2. there are additional macros for printing counter values with uppercase and lowercase Cyrillic letters (`\Asbuk` and `\asbuk` as analogs to `\Alph` and `\alph`).

We created an additional “shorthand” for Russian emdash (“---”): in our printing documents this sign is shorter a little, and it has spaces about 0.2 of font size (e.g. for size 10pt it is about 2pt) around and never breaks with word before emdash. *Note:* the macro for explicit hyphen sign (“-”) was, certainly, rewritten because of creating macro for Russian emdash.

4 Macros, which have some difficulties or questions in usage

In this section we discuss macros of breaking in text formulas. By Russian typography tradition we must repeat last sign in broken formula on the next line.

There is a package which includes macros for repeating signs in formulas.

This package includes two possible ways:

1. *hand breaking* – in this way $\binoppenalty = \relpenalty = 10000$; for breaking we must use commands $\brokenbin{ }$ and $\brokenrel{ }$;
2. *automatic breaking* – in this way $\relpenalty > \binoppenalty > 10000$; some signs are equal to $\mathcode=8000$ and divided into two groups: binary and relational signs $+ - < > =$ allow break after them, signs $* ([/ . ,$ protect break; also almost all mathematic signs are rewritten using new commands \brkbin and \brkrel to allow or protect breaking, e.g.:

```
\def\wedge {\brkbin{\mathchar"225E}}
\def\gg      {\brkrel{\mathchar"321D}}
\def\exists  {\mathchar"0239\unbrk}
\def\bigl#1 {\mathopen{\big#1}\unbrk}
\def\bigm#1 {\mathrel {\big#1}\unbrk}
\def\langle {\delimiter"426830A \unbrk}
```

the command \not , for example, must be redefined:

```
\def\not#1 {\brkrel{\mathchar"3236 #1}}
```

and so on. In this case \TeX breaks formulas by itself but sometimes we must handle breaking using special commands \unbrk or \allowbrk .

There are the drawbacks in this package but rather exotic:

- one must write $\$x \brkbin{\{+\}^1} y\$$ instead of $\$x +^1 y\$$;
- operators \sin must be written with arguments in parenthesis;
- in formulas like $x + \dots + y$ one must write $\$x \unbrk + \ldots + y\$$ to protect first breaking (or breaking signs must look forward);
- in case the signs \wedge and $_$ are redescribed, we cannot use $\mathcal{AMS}\text{-}\TeX$'s macros like $\Sb.\.\endSb$ (i.e. $\^{\bgroup}.\.\egroup$) otherwise it is impossible to use something as $\^{\leq}$ and $\^*$.

As you see we must rewrite all definitions for binary and relation signs and some signs which protect the break after them. In other words we must rewrite \TeX formats.

Now this style exists as additional² and is switched on by user.

² Now it is beta-version style.

5 Encoding and font problems

The file `russianb` was created for “nonlatin user” who uses Cyrillic alphabet. One of the problems is: there are a lot of different encodings in which Cyrillic letters have different codes.

Because of this, our file has some particular features as compared with Babel’s analogs.

To make this style independent of encoding, we have to use macros-names instead of Cyrillic letters themselves. Russian letters and signs in this style are used in macros for date (`\daterussian`) and the strings for four standard styles of L^AT_EX (`\captionrussian`), and also in commands for printing counter values with use of Cyrillic uppercase and lowercase letters (`\Asbuk` and `\asbuk`).

The macros-names for Cyrillic letters and some signs are switched on by a file-satellite (e.g. `lhrcod.sty`) which is created for necessary encoding. This file also switches on Cyrillic font family.³

In Russia, papers are typesetted by special encodings where the Cyrillic letters are joined with Latin ones and placed by necessary encoding in upper part of the code table, so there is the tradition of using Russian letters in macronames (the usage of familiar, Russian words is more convenient – Russian letters are letters too, aren’t they?). To tell more: there are packages which are based on Russian-word commands. So, in `russianb` Cyrillic letters were made as `\catcode\letter`.

Now we must say that `\mathcode` for Russian letters equals to 70??, i.e. we set class 7 – variable family and use font of `\fam0` (`\rm`) to join the Russian typography tradition.

6 What we must do

Now our style is adapted for Russian “8-bit” documentation, in which we sometimes use Latin fragments. In this case a file-satellite `russianb` (in current version we input `lhrcod` for Alternative encoding) simply declares new font family and encoding at the beginning of document⁴ and then toggles Russian/Latin hyphenation only – this way takes less memory. For Russian-Latin papers typesetted with transliteration (using Latin letters as Cyrillic) we must declare sometimes to Latin letters that they are Cyrillic, so we have to toggle fonts and encodings too. These two ways should be described and divided in `russianb`.

The main problem of file `russianb` is that we have to use macros-names of Cyrillic letters. Now these letters are described as `\def\CYRa{a}`. In this case the commands `\uppercase` and `\lowercase` don’t work correctly (e.g we can’t use standard style book

3. Now the switching on of the Cyrillic family is made for L^AT_EX_{2 ϵ} only.

4. Such file now is created for L^AT_EX_{2 ϵ} only.

– we will never get uppercase `\chaptername` in running heads).⁵ Maybe other definitions for letters or/and strings for styles will solve this problem.

The paragraphs above described some solutions of portability of the file `russianb` to different platforms. Since we have no practice in other platforms (e.g. Unix) we don't know about probable difficulties there.

Our work is the first attempt only and we hope to provoke discussion and further work of our colleagues. We hope also that `Omega` package can help to solve the described problems, especially those that are connected with encoding schemes and portability.

7 Acknowledgements

In conclusion we would like to note that a lot of our colleagues make the valuable contributions to creation of the file `russianb`: Mikhail Grinchuk, Eugenii Ivanov, Andrey Slepukhin, Yurii Tyumentsev, and others. We are very much obliged to all of them.

References

- [1] J. Braams. Babel, a multilingual style-option system for use with \LaTeX 's standard document styles. *TUGBoat*, 12.2:291–302, 1991.
- [2] P. Gilenson. *Spravochnik tekhnicheskogo redaktora*. Moscow, Kniga, 1979.
- [3] M. Goossens, F. Mittelbach, and A. Samarin. *The \LaTeX Companion*. Addison-Wesley, Reading, MA, 1994.
- [4] Y. Haralambous and J. Plaice. First application of `Omega`: Greek, arabic, khmer, poetica, iso 10646/unicode, etc. *TUGBoat*, 15.3:344–352, 1994.
- [5] A. Khodulev and I Makhovaya. *On \TeX experience in Mir Publishers*. Proceedings of the 7th Euro \TeX Conference, Prague, 1992.
- [6] O. Lapko. *MAKEFONT as part of CyrTUG-em \TeX package*. Proceedings of the eight European \TeX Conference, Gdańsk, 1994.

5. If Cyrillic letters would be described as `\def\CYRa{\char160}` or `\chardef \CYRa=160` we won't get uppercase/lowercase texts in principle.