
Kyrillisch für arme Leute: Setzen russischer Textpassagen mit LaTeX

Walter Schmidt
24. Mai 2005

`<http://home.vr-web.de/was/x/pmcyr.pdf>`

1 Die Aufgabe

In einem deutsch- oder englischsprachigen Text – oder jedenfalls in einem, der in lateinischer Schrift gesetzt ist – sollen einzelne russische Wörter oder ganze Absätze in kyrillischer Schrift erscheinen. Dabei wird eine normale »westliche« Tastatur verwendet. Das Kyrillische soll mit ASCII-Zeichen eingegeben werden, denn selbst wenn ein mehrsprachiger Eingabezeichensatz wie UTF-8 zur Verfügung steht, ist es unbequem, kyrillische Zeichen quasi »blind« von einer lateinisch beschrifteten Tastatur einzutippen.

Es ist offensichtlich, dass dazu bestimmte Voraussetzungen erfüllt sein müssen:

1. Man benötigt Fonts mit kyrillischem Zeichensatz, die im Aussehen zu der jeweils benutzten lateinischen Schrift passen. Also eine kyrillische Computer-Modern-Schrift, wenn der Text ansonsten mit CM gesetzt ist, oder eine kyrillische Times, wenn die Grundschrift Times ist, usw.
2. Ein Verfahren zur Eingabe der kyrillischen Zeichen als ASCII-Umschrift muss existieren. Es sollte einer halbwegs einleuchtenden Transliteration entsprechen; die Eingabe jedes einzelnen kyrillischen Buchstabens als Makro wäre natürlich nicht akzeptabel.
3. Das LaTeX-Format muss russische Trennmuster enthalten, damit innerhalb von russisch geschriebenen Absätzen die Silbentrennung funktioniert.
4. Das Babel-System wird zur Umschaltung zwischen den verschiedenen Trennmustern und Schriften gebraucht.

Ganz naiv könnte man vermuten, dass im »Cyrguide« [1] und in der Dokumentation von Babel [2] beschrieben ist, wie die Aufgabe zu bewältigen ist, doch leider helfen diese Dokumente nicht weiter: Sie nehmen erster Linie auf die Verwendung von kyrillischen Fonts mit dem sogenannten T2A-Zeichensatz Bezug, der aber nur zusammen mit *direkter* Eingabe der kyrillischen Zeichen sinnvoll verwendbar ist. Das ist zwar die bevorzugte Methode für ernsthaften kyrillischen Textsatz mit LaTeX, aber für »Gelegenheits-Schreiber« an einer westlichen Tastatur nicht unbedingt praktisch.

Wie russischer Textsatz unter den eingangs genannten Bedingungen mit LaTeX möglich ist, soll im Folgenden erklärt werden.

2 Kyrillische Fonts

Die oben genannten Voraussetzungen 1 und 2 werden erfüllt, wenn man kyrillische Fonts mit dem sogenannten OT2-Encoding verwendet. Dieses Encoding hat kyrillische Zeichen auf den Positionen, wo sich normalerweise ASCII-Zeichen befinden. Zusammen mit trickreich definierten Ligaturen wird eine halbwegs komfortable Eingabe allein unter Verwendung von ASCII-Zeichen ermöglicht. Der Nachteil ist allerdings, dass die automatische Silbentrennung damit nicht mehr perfekt funktioniert. Das liegt an den Eingabeligaturen und außerdem daran, dass OT2 nicht alle möglichen Kombinationen aus Buchstaben und diakritischen Zeichen »fertig« enthält. Letzteres Problem kennen wir auch vom OT1-Encoding. Von russischen Anwendern oder für den Satz komplett kyrillischer Texte wird OT2 normalerweise *nicht* verwendet. Für unseren Zweck ist es aber eine praktikable Alternative. Nachteilig ist – neben den Mängeln bei der Silbentrennung – dass es nur sehr wenige passende Fonts gibt:

- In praktisch jedem LaTeX-System stehen mit OT2-Encoding zwei Schriftfamilien namens `wncyr` und `wncyss` zur Verfügung, die sog. WNCYR-Fonts. Sie passen vom Stil her zu CM-Roman und CM-Sans, gelten aber – für russische Leser – als ästhetisch fragwürdig.
- Gut unterstützt wird die verbreitete Schriftkombination Times/Helvetica: Kyrillische Times- und Arial-Fonts im PostScript-Format sind als Bestandteil der Sammlung PSCyr¹ frei erhältlich. Virtuelle Fonts, die diese auf OT2-Encoding abbilden, sind ebenfalls verfügbar.²
- OT2-codierte Fonts der Familien Computer Modern Roman (`cmr`) und Sans (`cmss`) sind zwar vorhanden; diese sog. LH-Fonts existieren aber nur im METAFONT-Format, nicht als PostScript, so dass ihre Verwendung oft nicht in Frage kommt. Manche TeX-Installationen, z.B. VTeX, enthalten daher modifizierte Fontdefinitions-Dateien die den Schriftfamilien `cmr` und `cmss` fürs OT2-Encoding stattdessen die o. g. WNCYR-Fonts zuordnen; letztere sind nämlich auch im PostScript-Format vorhanden. (Die entsprechenden `fd`-Dateien sind in Abschnitt 7 dieses Dokuments abgedruckt, zusammen mit Hinweisen für ihre Installation.)

Um OT2-codierte Fonts überhaupt verwenden zu können, ist in der Präambel das Paket `fontenc` mit der Option `OT2` zu laden. Als *letzte* Paketooption *muss* das defaultmäßig im Text zu verwendende Encoding angegeben werden, also z.B. `T1` (oder `OT1`, falls man bei der Voreinstellung von LaTeX bleiben will):

```
\usepackage[OT2, T1]{fontenc}
```

3 Russische Trennmuster

Automatische Silbentrennung in russischen Textpassagen setzt geeignete Trennmusterdateien voraus.

Aus der Dokumentation des LaTeX-Systems sollte hervorgehen, ob russische Trennmusterdateien überhaupt vorhanden sind. Fehlen sie, so findet man ein gan-

1 [<ftp://ftp.vsu.ru/pub/tex/font-packs/pscyr/>](ftp://ftp.vsu.ru/pub/tex/font-packs/pscyr/)

2 [<http://home.vr-web.de/was/fonts.html>](http://home.vr-web.de/was/fonts.html)

zes Bündel von Dateien im CTAN-Verzeichnis `language/hyphenation/ruhyphen/`. Diese Dateien sind ins Verzeichnis `texmf/tex/generic/hyphenation/` des LaTeX-Systems – oder besser in ein Unterverzeichnis davon, etwa `russian/` – zu kopieren.

Die Trennmuster funktionieren nur für ein bestimmtes Schrift-Encoding, wobei T2A voreingestellt ist. Für OT2 müssen wir in einem geeigneten Verzeichnis (z. B. dort, wo sich auch andere lokale Konfigurationsdateien für LaTeX befinden) eine Datei namens `ruhyphen.cfg` anlegen, mit folgendem Inhalt:

```
\def\Encoding{ot2}
\input ruhyphen
```

Damit die russischen Trennmuster ins LaTeX-Format eingebaut werden, müssen sie in der Konfigurationsdatei für die Trennmuster aufgeführt sein. Die meisten LaTeX-Systeme verwenden eine Datei mit Namen `language.dat`. Der Eintrag in `language.dat` muss lauten:

```
russian      ruhyphen.cfg
```

Falls das LaTeX-System die Trennmuster nicht mittels `language.dat` festlegt, sondern in einer Datei namens `hyphen.cfg`, dann muss für die russischen Trennmuster darin folgender Eintrag stehen:

```
\newlanguage\l@russian \language=\l@russian
\begingroup
\def\Encoding{ot2}
\input ruhyphen
\endgroup
```

Die zusätzliche Datei `ruhyphen.cfg` wird in diesem Fall nicht gebraucht.

Damit eventuelle Änderungen bzgl. der Trennmuster auch tatsächlich wirksam werden, muss danach das LaTeX-Format neu erzeugt werden.

4 Texteingabe mit Babel

Wir wollen Babel verwenden, um Trennmuster und Fonts sprachabhängig umzuschalten. Das Paket muss mit der Option `russian` geladen werden, und als *letzte* Option ist die im Dokument als Voreinstellung gewünschte Sprache, z.B. `USenglish`, anzugeben. Wichtig ist, dass `babel` *nach* `fontenc` geladen wird: Wenn zu diesem Zeitpunkt das Encoding OT2 deklariert ist, dann »weiß« Babel, dass es dieses für die russischen Textteile verwenden soll, und nicht seine Voreinstellung T2A:

```
\usepackage[OT2, T1]{fontenc}
\usepackage[russian, USenglish]{babel}
```

Zum Setzen russischer Textteile innerhalb des Dokuments dienen dann der Befehl `\foreignlanguage` oder die Umgebung `otherlanguage*`. Z. B. ergibt

```
\foreignlanguage{russian}{rabota}
```

die Ausgabe `работа`, also das russische Wort für »Arbeit«.

Kyrillischer Text ist in der Umschrift einzugeben, die in Tabelle 1 zusammengefasst ist. Dort sind alle kyrillischen Zeichen aufgeführt, jeweils als Groß und Klein-

buchstaben, und danach die Eingabe. Wenn alternative Eingabeformen existieren, so ist diese ebenfalls angegeben. Die Tabelle enthält nicht nur die Zeichen des russischen Alphabets, sondern noch weitere, die für andere kyrillisch geschriebene Sprachen benötigt werden.

Tabelle 1: Eingabe kyrillischer Zeichen mittels Washington-Cyrillic-Umschrift.

A	a	A	a			O	o	O	o		
Б	б	B	b			П	п	P	p		
В	в	V	v			Р	р	R	r		
Г	г	G	g			С	с	S	s		
Д	д	D	d			Т	т	T	t		
Ђ	ђ	Dj	dj	D1	d1	Ѓ	ѓ	C1	c1		
Ѓ	ѓ	\'G	\'g			Ќ	ќ	\'K	\'k		
Е	е	E	e			У	у	U	u		
Ё	ё	\"E	\"e	E0	e0	Ф	ф	F	f		
Є	є	E2	e2			Х	х	Kh	kh	H	h
Ж	ж	Zh	zh	Z1	z1	Ц	ц	Ts	ts	C	c
З	з	Z	z			Ч	ч	Ch	ch	Q	q
И	и	I	i			Ї	ї	D2	d2		
І	і	\CYRII	\.{i}	I1	i1	Ш	ш	Sh	sh	X	x
Й	й	\"I	\"i			Щ	щ	Shch	shch	W	w
Ј	ј	J	j			Ъ	ъ	P2	p2		
Ў	ў	\U{I}	\U{i}	IO	io	Ы	ы	Y	y		
К	к	K	k			Ь	ь	P1	p1		
Л	л	L	l			Э	э	E1	e1		
Љ	љ	Lj	lj	L1	l1	Ю	ю	Yu	yu	J2	j2
М	м	M	m			Я	я	Ya	ya	J1	j1
Н	н	N	n			С	с	D3	d3		
Њ	њ	Nj	nj	N1	n1	«	»	<	>		
№		NO									

Zusätzlich stehen einige vom Babel-Paket definierte Befehle zur Verfügung, die in dessen Dokumentation [2] im Abschnitt »The Russian language« erklärt sind. Besonders zu achten ist auf die eventuell nötige Trennung der Eingabeligatur t-s. Wo die kyrillischen Buchstaben т und с tatsächlich aufeinanderfolgen, wie im Wort советский, ist die Ligatur mittels "| zu verhindern: sovetski|o.

Natürlich sind mit OT2-codierten Schriften auch alle übrigen Features des Babel-Systems verfügbar, einschließlich der Umschaltung der Schlüsselwörter (wie z.B. Inhaltsverzeichnis → Содержание). Für längere russischsprachige Abschnitte oder komplette Dokumente wird das OT2-Encoding aber nicht empfohlen.

5 Es geht auch ohne Babel

Notfalls lässt sich das OT2-Encoding auch ohne Hilfe von Babel und ohne Silbentrennung nutzen. Das kann beispielsweise dann Sinn machen, wenn man die Trennmuster nicht auf OT2-Encoding umstellen kann, oder wenn gar keine Trennmuster für die jeweilige Sprache vorhanden sind – nicht nur das Russische wird ja mit kyrillischer Schrift geschrieben!

Sinnvollerweise definiert man sich ein Makro, das für sein Argument auf OT2-Encoding umschaltet und die Silbentrennung unterdrückt, damit nicht etwa nach den Regeln der »umgebenden« Sprache getrennt wird:

```
\newcommand{\textcyr}[1]{%
  {\donothyphenate\fontencoding{OT2}\selectfont #1}}
```

Das Makro `\donothyphenate` müssen wir uns folgendermaßen definieren:

```
\newlanguage\nohyphens
\newcommand{\donothyphenate}{%
  \language=\nohyphens}
```

Voraussetzung ist ferner, dass das OT2-Encoding zuvor mittels `fontenc` deklariert wurde, wie oben beschrieben. Dann erzeugt z.B. die Eingabe von `\textcyr{chai0}` das russische Wort für »Tee«: чай.

Alle Eingabe-Tricks gemäß Tabelle 1 sind auch ohne Babel möglich. Ligaturen müssen mittels `{\kern0pt}` verhindert werden, falls nicht der Befehl `|` beispielsweise von `ngerman.sty` bereitgestellt wird.

6 Wenn die passenden Fonts fehlen. . .

Mit dem beschriebenen Vorgehen erleidet man Schiffbruch, wenn die aktuell verwendete Schriftfamilie nicht mit OT2-Encoding zur Verfügung steht. Als Rückfallebene wählt LaTeX dann CM-Roman und ignoriert Auszeichnungen wie kursiv oder fett. Nachdem das normalerweise so nicht erwünscht ist, muss man die Umschaltung auf »Russisch« mit der expliziten Auswahl einer Schriftfamilie verbinden, die mit OT2-Encoding vorhanden und so gut wie möglich mit der lateinischen kompatibel ist. Um beispielsweise die kyrillische Times (`ptm`) auswählen, muss man schreiben:

```
{\fontfamily{ptm}\selectfont%
  \foreignlanguage{russian}{kyrillischer Text... }}
```

Auch das im vorigen Abschnitt vorgeschlagene Makro `\textcyr` kann um die Auswahl einer bestimmten Schriftfamilie ergänzt werden.

Die Situation ist unbefriedigend, aber letztlich auch nicht anders als bei den phonetischen Schriften (TIPA-Fonts), die auch nur im Stil von CM-Roman, CM-Sans, Times und Helvetica vorhanden sind. Es ist zu hoffen, dass in Zukunft noch weitere kyrillische Fonts – auch solche, die nicht speziell für TeX entworfen wurden – mit OT2-Encoding nutzbar gemacht werden.

7 fd-Dateien für die WNCYR-Fonts als Computer-Modern-Ersatz

Die im Folgenden abgedruckten Dateien `ot2cmr.fd` und `ot2cmss.fd` deklarieren die WNCYR-Fonts als Ersatz der nicht im Type1-Format vorhandenen kyrillischen LH-Fonts, siehe Abschnitt 2. Damit LaTeX sie verwendet, kann man sie in einem Verzeichnis installieren, wo sie *vor* den entsprechenden standardmäßigen `fd`-Dateien gefunden werden. Die meisten TeX-Systeme besitzen einen sog. »lokalen« Dateibaum, wo man z. .B. eigene Makropakete ablegt; dort ist ein geeigneter Platz.

```

%% File ot2cmr.fd
%%
\ProvidesFile{ot2cmr.fd}
\DeclareFontFamily{OT2}{cmr}{\hyphenchar\font45 }
\DeclareFontShape{OT2}{cmr}{m}{n}{<->wncyr10}{}
\DeclareFontShape{OT2}{cmr}{m}{it}{<->wncyi10}{}
\DeclareFontShape{OT2}{cmr}{m}{sc}{<->wncysc10}{}
\DeclareFontShape{OT2}{cmr}{b}{n}{<->wncyb10}{}
\DeclareFontShape{OT2}{cmr}{bx}{n}{<->ssub*cmr/b/n}{}
\endinput
%%
%% End of file ot2cmr.fd

%% File ot2cmss.fd
%%
\ProvidesFile{ot2cmss.fd}
\DeclareFontFamily{OT2}{cmss}{\hyphenchar\font45 }
\DeclareFontShape{OT2}{cmss}{m}{n}{<->wncyss10}{}
\endinput
%%
%% End of file ot2cmss.fd

```

Literatur

- [1] Vladimir Volovich, Werner Lemberg, LaTeX3 Project Team: *Cyrillic language support in LaTeX*. Bestandteil der Online -Dokumentation von LaTeX, Datei cyrguide.dvi oder .pdf.
- [2] Johannes Braams: *Babel, a multilingual package for use with LaTeX's standard document classes*. Bestandteil der Online-Dokumentation von LaTeX, Datei user.dvi oder .pdf.
- [3] Holm Sieber: *Russischer Textsatz mit LaTeX*. Die TeXnische Komödie 9(1997), H.2