

Reviving the Traditional Russian Orthography for the 21st Century

Sergei Winitzki

Text By The Bay 2015

April 24, 2015

Old Russian Orthography: what and why

- The Russian orthography was reformed in 1918
 - ▶ 4 letters Ъ, Ѣ, Іі, Vv were physically removed from printing presses
 - ★ Often, Ъ was also removed by mistake
- Classic Russian literature used the old orthography:
 - ▶ Pushkin, Turgenev, Dostoyevsky, Tolstoy, Chekhov, Nabokov, ...

Old English orthography: example

Hobbes, *De cive* (1651)

I. **A**LL Authors agree not concerning the definition *of the Naturall Law*, who notwithstanding doe very often make use of this terme in their Writings. The Method therefore, wherein we begin from definitions, and exclusion of all equivocation, is only proper to them who leave no place for contrary Disputes; for the rest, if any man say, that somewhat is done against the Law of Nature, one proves it hence, because it was done against the generall Agreement of all the most wise, and learned Nations: But this declares not who shall be the judg of the wisdome and learning of all Nations:

Old Russian orthography: example

Nabokov, *Storm* (1930)

Г Р О З А

На углу, подъ шатромъ цвѣтущей липы, обдало меня буйнымъ благоуханіемъ. Туманныя громады поднимались по ночному небу, и когда поглощенъ былъ послѣдній звѣздный просвѣтъ, слѣпой вѣтеръ, закрывъ лицо рукавами, низко пронесся вдоль опустѣвшей улицы. Въ тусклой темнотѣ, надъ желѣзнымъ ставнемъ парикмахерской, маятникомъ заходилъ висячій щитъ, золотое блюдо.

Вернувшись домой, я засталъ вѣтеръ уже въ комнатѣ: — онъ хлопнулъ оконной рамой и поспѣшно от-

Main challenges in processing old Russian orthography

- As of 2000, no 8-bit encodings contained the letters Ъъ, Ѡѡ, Ѣѣ, Ѥѥ
 - ▶ Unicode support was not widespread
- No keyboard layouts, no screen fonts, few vector fonts
 - ▶ L^AT_EX font support: nonstandard font encoding by AMS
- No spelling checker support, no OCR, no hyphenation

State of the art in 2015

- The letters Ъ, Ѣ, Ѡ, ѡ, Ѣ, ѣ, Ѡ, ѡ are present in most Cyrillic fonts
 - ▶ typesetting, printing supported through Unicode
 - ▶ L^AT_EX font support: OT2 font encoding only!
- Some spelling checker support (ispell, AbiWord, Open Office)
- Some language support for OCR
- No hyphenation support, no standard keyboard layouts

The KOI8-C project (1999-2003)

- <http://oldrus-ispell.sf.net>
- KOI8-C encoding ([IETF draft](#))
- Bitmapped X-Window fonts (BDF), the `xcyr` package
 - ▶ Became part of `xfonts-cyrillic`
- `ispell` dictionary for old Russian orthography
- PostScript typesetter using BDF as Type 3 fonts
- A primitive converter from new to old orthography

The KOI8-C encoding and screen fonts in Netscape

Viewing Nabokov, *Despair* (1934)

росту не знаю, съ чего начать. Смѣшонъ пожилой человѣкъ, который бѣгомъ, съ прыгающимъ, догналъ {5} послѣдній автобусъ, но боится вскочить на ходу, и виновато улыбаясь, еще нужно не смѣю вскочить? Онъ воетъ, онъ ускоряетъ ходъ, онъ сейчасъ уйдетъ за уголъ, непонятно изъ этого рассказа. Образъ довольно громоздкій. Я все еще бѣгу.

Моя мать была ревельскій нѣмецъ, по образованію агрономъ, покойная мать -- чисто-русская. Съ детства, въ жаркіе лѣтніе дни она, бывало, въ сиреневыхъ шелкахъ, томная, съ вѣеромъ въ рукѣ, ползала по кушеткамъ, кушала шоколадъ, и наливались сѣнокоснымъ вѣтромъ лиловые паруса спущенныхъ штормовыхъ парусовъ. Этого подданного, интернировали, -- я только-что поступилъ въ петербургскій университетъ, и в четырнадцатого до середины девятнадцатая года я прочелъ тысяча восемнадцать книгъ, -- великому я на три мѣсяца застрялъ въ Москвѣ и тамъ женился. Съ двадцатого года проживалъ въ Петербургѣ, уже переваливъ лично за тридцать пять --

дополненіе: насчетъ матери я совралъ. По настоящему она была дочь мелкаго мѣщанина, -- про которую я не скажу ей.

koi8c-table.gif (671x312) 0.0s 100% [1/1] b0/c0/g0															
80	81	82	83	84	85	86	87	88	89	8A	8B	8C	8D	8E	8F
Ъ	Г	А	І	В	...	Ѡ	ѡ	Ѣ	Ѥ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ
90	91	92	93	94	95	96	97	98	99	9A	9B	9C	9D	9E	9F
Ѣ	Ѥ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ѹ	Ѻ	Ѽ	Ѿ	Ѻ	Ѽ	Ѿ	Ѻ
AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP
	Ѣ	Ѥ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ѹ	Ѻ	Ѽ	Ѿ	Ѻ	Ѽ	Ѿ
BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP
	Ѣ	Ѥ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ѹ	Ѻ	Ѽ	Ѿ	Ѻ	Ѽ	Ѿ
CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP
	Ѣ	Ѥ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ѹ	Ѻ	Ѽ	Ѿ	Ѻ	Ѽ	Ѿ
DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP
	Ѣ	Ѥ	Ѭ	Ѯ	Ѱ	Ѳ	Ѵ	Ѷ	Ѹ	Ѻ	Ѽ	Ѿ	Ѻ	Ѽ	Ѿ

Working with the oldrus-ispell dictionary

Proofreading Melgunov's "Red Terror in Russia" (1924)

The screenshot shows a text editor window with a menu bar (File, Edit, Search, Preferences, Shell, Macro, Windows) and a status bar (Shell Command in Progress -- Press Ctrl+. to Cancel). The main text area contains a passage from Melgunov's "Red Terror in Russia" (1924) in Old Russian Cyrillic. The text is as follows:

"Еженедѣльникъ Ч. К." *(6) никогда ничего больше {39} не было опубликовано. А между тѣм мы знаем, что людей в эти дни в Москвѣ по общим свѣдѣніям было разстрѣлено больше 300 *(7).
Тѣ, которые сидѣли в эти поистинѣ мучительные дни в Бутырской тюрьмѣ, когда были арестованы тысячи людей из самых разнообразных общественных слоев, никогда не забудут своих душевных переживаній. Это было время
террора".
присутство
автомобили
каждом авт
"с вещами"
связывать
сидѣл в эт
кошмары. В
"В пам
разстрѣл и

A vertical scrollbar is visible on the left side of the text area. A small window titled "ispell" is open, showing the word "без" (bez) and its suggestions: "0: бей", "1: безе", "2: безо", "3: безъ". The word "без" is highlighted in the text area.

Main challenges

Main challenges

Main challenges