

*The*

TRANSLITERATOR

*for ConT<sub>E</sub>Xt*

MANUAL

---

The *Transliterat*or module and mini-manual,  
by Philipp Gesang, Radebeul.  
Mail any patches or suggestions to

philipp -dot- gesang -at- alumni -dot- uni-heidelberg -dot- de  
<https://phi-gamma.net>

# CONTENTS

USAGE AND FUNCTIONALITY .....	2
<i>Overview</i> .....	2
<i>Loading and Configuring the Module</i> .....	2
INTRODUCTION .....	5
EXAMPLES .....	8
<i>Cyrillic scripts</i> .....	8
<i>iso 9 and derivatives</i> .....	8
<i>“Scientific” transliteration</i> .....	10
<i>Serbian</i> .....	11
<i>Bulgarian</i> .....	12
<i>Legacy national transcriptions</i> .....	12
<i>Glagolitic</i> .....	14
<i>Greek</i> .....	14
REFERENCES .....	15

# USAGE AND FUNCTIONALITY

## OVERVIEW

The Transliterator provides two commands: `\setuptransliterator` preferably goes into the preamble and allows for global configuration. The Transliterator is invoked locally by `\transliterate` which does the actual transliteration of text passages.

```
\setuptransliterator [...*...]
```

* mode	=	<code>ru_old</code>	<code>ru</code>	<code>ru_transcript_de</code>	<code>ru_transcript_de_exp</code>	<code>ru_transcript_en</code>	<code>all</code>	<code>sr_tolt</code>	<code>sr_tocy</code>	<code>bg_de</code>	<code>iso9_ocs</code>	<code>ocs</code>	<code>ocs_gla</code>	<code>ru_cz</code>	<code>ocs_cz</code>	<code>gr</code>	<code>gr_n</code>
hyphenate	=	<code>cs</code>	<code>sk</code>	<code>hr</code>													
deficient_font	=	<code>yes</code>	<code>no</code>														
hinting	=	<code>yes</code>	<code>no</code>														
sr_exceptions	=	<code>yes</code>	<code>no</code>														

```
\transliterate [...*...]
```

\* inherits: `\setuptransliterator`<sup>OPT</sup>

## LOADING AND CONFIGURING THE MODULE

In order to use the Transliterator in a document we put the following somewhere before `\starttext`.

```
\usemodule[transliterator]
```

Although it has some defaults already set at this point they will most likely not correspond to what is needed in the document. To override the presets we use the command `\setuptransliterator[#1]`. It takes a comma separated list of two key-value pairs: `mode` and `hyphenate`. Through `mode` we specify the transliteration method. By the time of this writing this can be one of the following set:

*Nota bene:* The description at this point only serves as a placeholder as the transliteration modes are discussed in detail later in this document.

mode	description
<code>all</code>	ISO 9 complete
<code>bg_de</code>	Bulgarian, German „scientific“ transliteration
<code>gr</code>	transliteration for Greek
<code>gr_n</code>	transliteration for Greek obeying nasalizations
<code>iso9_ocs</code>	== <code>all</code> plus non-ISO additions for Old (Church) Slavonic
<code>ocs</code>	“scientific” transliteration for Old (Church) Slavonic
<code>ocs_cz</code>	Czech transcription for Old (Church) Slavonic
<code>ocs_gla</code>	“scientific” transliteration for Old (Church) Slavonic / Glagolitic alphabet
<code>ru</code>	ISO 9 Russian
<code>ru_cz</code>	Czech transcription for Russian
<code>ru_old</code>	ISO 9 Russian plus pre-1918 chars (the default)
<code>ru_transcript_de</code>	German transcription for Russian
<code>ru_transcript_en</code>	English transcription for Russian
<code>sr_tocy</code>	Serbian, Latin to Cyrillic
<code>sr_tolt</code>	Serbian, Cyrillic to Latin

**Table 1** Transliteration modes.

Through the `hyphenate` argument it is possible to adjust the language that is used for hyphenation. Specifying `\setuptransliterator[hyphenate=nl]` will let every transliterated part of the document be processed according to dutch rules, leaving the overall `\language[#1]` configuration unchanged for the rest of the content.

Another argument, `deficient_font` can be used in combination with the modes `all`, `ru_old` and `iso9_ocs`. It lets you circumvent the deficiency that some fonts show concerning the characters that ISO 9 assigns to cyrillic “Ь” and “ь”. Set it to `true` to enable it.

The actual transliteration is done using the macro `\transliterate[#1]{#2}`. The second argument takes the raw string in the original language that we want to process, while the first, optional argument accepts local adjustments for `mode` and `hyphenate`. Thus, we would typeset one of Epicuro’s sayings like this:

```
\transliterate[mode=gr]{κακὸν ἀνάγκη, ἀλλ’ οὐδεμία ἀνάγκη ζῆν
μετὰ ἀνάγκης}
```

which yields “kakon anagkē, all’ udemia anagkē zēn meta anagkēs” in the PDF output. Alternatively there is an environment, `\startttransliterate[#1]`, as well, that takes the same arguments.

There are two special switches for the *Serbian* patterns, `hinting` and `sr_exceptions`, allowing for a little more fine-tuning. If activated, `hinting`

provides the special character “\*” as a means to indicate positions, where the sequences “lj” and “nj” are to be treated as separate consonants. E. g. `\transliterate[mode=sr_tocy]{in*jekcija}` is correctly transliterated as инјекција, and not ињекција. Likewise, further exceptions that are internally represented as a lookup table can be toggled off or on by the `sr_exceptions` switch. This pertains to words like “nadživeti” (result: надживети) but may lead to accidental false positives in cases that the module author didn’t foresee. By default both hinting and lexical exceptions are set to `yes`.

For orientation purposes the Transliterator comes with two macros that allow for closer inspection of the internal tables. `\showOneTranslitTab{#1}` outputs, obviously, a single table; their identifiers can be found in the `trans_tables_*.lua` files in the transliterator directory. The lazy alternative is `\showTranslitTabs` which prints all registered tables in a row nicely formatted as indexable sections. (Be warned, this may take some time.)

## INTRODUCTION

*What's all this, then?*  
GRAHAM CHAPMAN

At the first glance, *transliteration* – the accurate representation of letters from one alphabet in another – seems obsolete after the advent of Unicode which made its way even into T<sub>E</sub>X lately. Why not just go on and write down everything in the original script? But still there are lots of situations where transliteration is desirable, e. g. some scholarly habits might prescribe it in the main text with citations in footnotes left in the original alphabet; or transliteration might alleviate comparison within one language that happens to be written in different scripts; finally, including text in a foreign script might be impossible if there is no appropriate font which fits the main text. However, it is still most convenient for the writer to keep the untransliterated original in the document source as this allows for reusing it in another context where different transliterations rules might apply. The Transliterator module is meant to provide both: have the original in the source and a transliteration only in the final document.

Another way of handling foreign languages is *transcription*. It aims at producing some representation that does not rely on symbolisms alien to the language and thus to be at least “pronouncable” without further knowledge. As transcription methods are language specific and highly idiosyncratic they complicate the restoration of the original phrase because information may be lost. The Transliterator provides means of transcription as well but in most cases you should refrain from using them (`[mode=ru_transcript_en]`, `[mode=ru_transcript_de]`).

For Cyrillic scripts the best quality is achieved using the standardized transliteration according to ISO 9.<sup>1</sup> This method not only covers all contemporary languages that are written in a variety of Cyrillic but provides a bijective mapping on latin characters as well. Consequently, you can unambiguously revert the transliteration into its original form which was impossible with previous versions of ISO 9 because they contained several exceptions depending on the original language. Although fifteen years old it has not yet made its way

---

<sup>1</sup> ISO (1995).





there is no simple way to impose order onto those dictionaries the rules for one transliteration method are, if needed, distributed over more than one table which will be applied successively to ensure that multi-character rules are processed first.

mode	time(1) in s	ConT <sub>E</sub> Xt
<none>	8.98	8.82
all	8.37	8.25
ru_cz	8.61	8.48
ru_transcript_en	9.26	9.10
ru_transcript_de	14.83	14.71

**Table 1** Processing time for corpus Evgenij Onegin according to GNU time(1) and the ConT<sub>E</sub>Xt stats.

Following suggestions from the mailing list, the Transliterator uses *LPeg* when substituting. This means a huge speed improvement for most substitution modes when compared to the older mechanism that used `string.gsub` iteratively. In ordinary use when transliterating single words or short phrases the Translit-

erator should have little impact on document processing time at large, with the exception of the German transcription mode, perhaps.<sup>5</sup> Transliterating (and typesetting in MKIV) Aleksandr Puškin’s verse novel Evgenij Onegin, a corpus of about 27000 words, in `[mode=all]` shows little to no delay at all. In fact, typesetting cyrillic letters with russian hyphenation seems slow things down so much that transliteration may be faster and uses slightly less memory.<sup>6</sup>

<sup>5</sup> The problem lies within the rule set for the German transcription which dictates different instructions depending on the environment of a character; these may conflict, i. e. it is impossible to substitute a character stream in a single run as some rules may apply only to the result of previous rule. Let me know if there’s a way to tell LPeg to backtrack to the last character of a match and not to continue on the next.

<sup>6</sup> On an IBM T43: 2.6.32-ARCH #1 SMP PREEMPT Tue Feb 9 14:46:08 UTC 2010 i686 Intel(R) Pentium(R) M processor 1.60GHz GenuineIntel GNU/Linux.

## EXAMPLES

### CYRILLIC SCRIPTS

#### ISO 9 AND DERIVATIVES

Several transliteration rules are either strictly ISO 9 compliant (`ru`, `ru_old`, `all`) or contain ISO 9 as a subset (`iso9_ocs`).<sup>7</sup>

В ворота гостиницы губернского города NN въехала довольно красивая рессорная небольшая бричка, в какой ездят холостяки: отставные подполковники, штабс-капитаны, помещики, имеющие около сотни душ крестьян, — словом, все те, которых называют господами средней руки. В бричке сидел господин, не красавец, но и не дурной наружности, ни слишком толст, ни слишком тонок; нельзя сказать, чтобы стар, однако ж и не так чтобы слишком молод.

V vorota gostinicy gubernskogo goroda NN v"ehala dovol'no krasivaâ ressornaâ nebol'shâa brička, v kakoj ezdât holostâki: otstavnye podpolkovniki, štabs-kapitany, pomešiki, imeûšie okolo sotni duš krest'ân, — slovom, vse te, kotoryh nazyvaût gospodami srednej ruki. V bričke sidel gospodin, ne krasavec, no i ne durnoj naružnosti, ni sliškom tolst, ni sliškom tonok; nel'zâ skazat', čtoby star, odnako Ź i ne tak čtoby sliškom molod.

**Figure 1** [`mode=ru,hyphenate=cs`] *Transliteration rules for the contemporary russian alphabet.*

А сведется віра, убьють сотцкого в селѣ, ино тебѣ взяти полтина, а не сотцкого, ино четырь гривны, а намъ віръ не тати в Новѣгородѣ; а о убиствѣ віръ нѣтъ. А что волости, честны король, новгородскіе, ино тебѣ не держати своими мужи, а держати мужми новгородскими. А что пошлина в Торжку и на Волоцѣ, тивунъ свои держати на своei чясти, а Новугороду на своei чясти посадника держати. А се волости новгородскіе: Волокъ со всѣми волостми, Торжокъ, Бѣжціи, Горodeцъ Палецъ, Шпининъ, Мелеця, Егна, Заволочье, Тирь, Пермь, Печера, Югра, Вологда с волостми.

A svedetsâ vira, ub'ût" sotckogo v selě, ino tebě vzâti poltina, a ne sotckogo, ino četyrě grivny, a nam" vir" ne taiti v Nověgorodě; a o ubistvě vir" nět". A čto volostî, čestny korol', novgorodckie, ino tebě ne deržati svoimi mužî, a deržati mužmi novgorodckimi. A čto pošlina v Toržku i na Volocě, tivun" svoi deržati na svoei čâstî, a Novugorodu na svoei čâstî posadnika deržati. A se volosti novgorodckie: Volok" so vsěmi volostmi, Toržok", Běžci, Gorodec' Palec', Šipin", Melecâ, Eгна, Zavoloče, Tir", Perm', Pečera, Ūgra, Vologda s volostmi.

**Figure 2** [`mode=ru_old,hyphenate=cs`] *With additional characters for pre-1981 Russian orthography (100 per cent iso 9).*

<sup>7</sup> Unfortunately ConTeXt still lacks language files for some of them so please excuse the inadequate hyphenation in these cases.

Беларуская мова, мова беларусаў, уваходзіць у сям'ю індаеўрапейскіх моў, яе славянскай групы і ўсходнеславянскіх моваў падгрупы, на якой размаўляюць у Беларусі і па ўсім свеце, галоўным чынам у Расіі, Украіне, Польшчы. Б. м. падзяляе шмат граматычных і лексічных уласцівасцяў з іншымі ўсходнеславянскімі мовамі (гл. таксама: Іншыя назвы беларускай мовы і Узаемныя ўплывы усходнеславянскіх моваў).

Belaruskaà mova, mova belarusaŭ, уваходзіць у сям'ю індаеўрапейскіх моў, яе славянскай групы і ўсходнеславянскіх моваў падгрупы, на якой размаўляюць у Беларусі і па ўсім свеце, галоўным чынам у Расіі, Украіне, Польшчы. Б. м. падзяляе шмат граматычных і лексічных уласцівасцяў з іншымі ўсходнеславянскімі мовамі (гл. таксама: Іншыя назвы беларускай мовы і Узаемныя ўплывы ўсходнеславянскіх моваў).

**Figure 3** [mode=all,hyphenate=cs] *The complete cyrillic mapping from iso g; transliterating Belarusian.*

Українська мова (застарілі назви – руська мова, проста мова [...]) – слов'янська мова, державна в Україні та одна з трьох «офіційних мов на рівних засадах» у невизнаній Придністровській Молдавській Республіці. За різними оцінками загалом у світі українською мовою говорить від 41 млн. до 45 млн. осіб, вона входить до третього десятка найпоширеніших мов світу.

Ukraïns'ka mova (zastarilii nazvi – rus'ka mova, prosta mova [...]) – slov'jans'ka mova, deržavna v Ukraïni ta odna z tr'oh «oficijnih mov na rïvnih zasadah» u neviznanij Pridnistrovs'kij Moldavs'kij Respublïci. Za rïznimi ocïnkami zagalom u svïti Ukraïns'koŭi movoŭi govorit' vid 41 mln. do 45 mln. osïb, vona vhodit' do tret'ogo desätka najpošïreniših mov svïtu.

**Figure 4** [mode=all,hyphenate=cs] *The complete cyrillic mapping from iso g; transliterating Ukrainian.*

Srpski jezik je jedan od slovenskih jezika iz porodice indoevropskih jezika. Prvi pisani spomenici u srpskoj redakciji staroslovenskog jezika potiču iz XI i XII veka. Srpski jezik je standardni jezik u službenoj upotrebi u Srbiji, Bosni i Hercegovini i Crnoj Gori, a u upotrebi je i u drugim zemljama gdje žive Srbi, među ostalima i u Hrvatskoj.

Srpski jezik je jedan od slovenskih jezika iz porodice indoevropskih jezika. Prvi pisani spomenici u srpskoj redakciji staroslovenskog jezika potiču iz XI i XII veka. Srpski jezik je standardni jezik u službenoj upotrebi u Srbiji, Bosni i Hercegovini i Crnoj Gori, a u upotrebi je i u drugim zemljama gdje žive Srbi, među ostalima i u Hrvatskoj.

**Figure 5** [mode=all,hyphenate=cs] *The complete cyrillic mapping from iso g; transliterating Serbian.*

ЧѢТО СѦ ДѢЈЕТЕ ПО ВѢРЕМ'НЕМЪ : ТО ВЪИДЕТО ПО ВѢРЬ-  
 МНЕМЪ : ПРИКАЗАНО БУДЕТЕ ДОБРЫМЪ ЛЮД'ВМЪ :  
 А ЛЮБО ГРАМОТОЮ ОУТВѢРДАТЕ : КАКО ТО БУДЕТЕ  
 ВСЕМЪ ВѢДОМЪ : ИЛИ КТО ПОСЛА ЖИВЫИ ОСТАНИТЕ  
 СѦ : ТОГО ЛѢТ' КОЛИ АЛЪБРАХЪТЪ : ВЛДКА РИЗКИИ  
 ОУМЪРАЪ : ОУЗДОУМАЪЪ КНАЗЪ СМОЛ'НЕСКИИ : МЪСТИ-  
 САДЪЪ : ДВАДЪЪ СЪЪ : ПРИСАЛАЪ ВЪ РИГЮ СВОЕГО ЛУЧЪ-  
 ШЕГО ПОПА : ИЕРЬМЕА : И СЪ НИМЪ ОУМНА МУЖА ПАН-  
 ТЕЛЯ : ИСВОЕГО ГОРДА СМОЛ'НЕСКА : ТА ДВА БЫЛА ПО-  
 САЪЪ ОУ РИЗЪ : ИЗ РИГЪ ИХАЛИ НА ГОЧЬКИИ ВЕРЬГО :  
 ТАМО ТВЕРДИТИ МИРЪ :

Čto se dějete po věrem'nem' : to otideto  
 po věr'm'nem' : prikazano budete dobrým'  
 lúďem' : a lúbo gramotoú utvĕrďet' : kako  
 to budete všem' vědom' : ili kto posl' živyyi  
 ostanĕt' se : togo lĕt' koli al'braht' : vldka  
 rizkii um'rl' : uzdumal' knežĕ smol'neskii :  
 m'stislav' : dvďv' sm' : prislal' v' rigu svo-  
 jeho lučšego popa : jer'meà : i s' nim' um'na  
 muža pantel'ã : isvojego gorda smol'neska :  
 ta dva byla posl'm' u rizĕ : iz rigy jehali na  
 gočkyyi ber'go : tamo tverditii mir' :

**Figure 6** [mode=iso9\_ocs,hyphenate=cs] *Transliteration rules accord-  
ing to iso 9 with additions for Old (Church) Slavonic.*

#### “SCIENTIFIC” TRANSLITERATION

These transliterations are widely used among scholars, mainly linguists and,  
 to a lesser extent, historians. They comprise large character sets in order to  
 represent the original text adequately and facilitate comparison of texts of  
 the same language written in different scripts; they are not, however, as easily  
 reversible as ISO 9.

Се научемъ повѣсть сию. По потопѣ . перене снве  
 Ноеви . раздѣлиша землю . Симъ . Хамъ . Афетъ .  
 и гаса възтокъ . Симвови Персида . Ватръ . тоже и до  
 Индикиа в долготу и в ширину [и до Ниркокуриа]  
 такжеже рещи ѿ възтока и до полуденья . и Сурина  
 . и Индиа по Ефратъ рѣку . Вавилонъ . Кордунъ  
 . Асуране . Мисопотамира . Аравиа . старѣвшима .  
 Елмансъ . Инди . Равина . на вса Д.

Se načnem' povĕšť siju. Po potopĕ . pervie  
 snve Noevi . razdĕliša zemljū . Sim' . Ham' .  
 Afet' . i jase' v'vstok' . Simovi Persida .  
 Vatr' . tože i do Indikija v' dolgotu i v' širinu  
 [i do Nirokuria] jakože rešti ot' v'vstoka i do  
 poludĕnja . i Surija . i India po Efrat' rĕku  
 . Vavilon' . Korduna . Asurene . Misopota-  
 mira . Aravija . starĕjšaja . Elmais' . Indi .  
 Ravija . na vsĕ D.

**Figure 7** [mode=ocs,hyphenate=cs] *Transliteration for Old Slavonic  
used in Slavic studies, taken from the excellent book of Birnbaum/Schaeken  
(1999).<sup>8</sup>*

<sup>8</sup> This one and both of the following Czech transliterations, although elegantly dealing with hard and weak signs by taking characters from the Cyrillic alphabet, are not unquestioned from a typographical point of view: “If contrasting faces are used for phonetic transcriptions and main text, each entire phonetic word or passage, not just the individual phonetic characters, should be set in the chosen phonetic face. Patchwork typography, in which the letters of a single word come from different faces and fonts, is a sign of typographic failure. [...] Such mixtures are almost sure to fail unless all the fonts involved have been designed as a single family.” (Bringhurst (2008)) From this follows that it is advisably to reconsider your font whether it indeed provides the needed glyphs from Russian as well.

Прошло семь лет после 12-го года. Взволнованное историческое море Европы улеглось в свои берега. Оно казалось затихшим; но таинственные силы, двигающие человечество (таинственные потому, что законы, определяющие их движение, неизвестны нам), продолжали свое действие. Несмотря на то, что поверхность исторического моря казалась неподвижною, так же непрерывно, как движение времени, двигалось человечество. Слагались, разлагались различные группы людских сцеплений; подготавливались причины образования и разложения государств, перемещений народов.

Prošlo sedm let posle 12-go goda. Vzvolnovanoo istoričeskoo more Evropy ulegloš v svoi berega. Ono kazaloš zatičšim; no ta-instvennye sily, dvigajuščie čelovečstvo (ta-instvennye potomu, čto zakony, opredeljajuščie ich dviženie, neizvestny nam), prodolžali svoje dejstvie. Nesmotrja na to, čto poverchnost' istoričeskogo morja kazalaš nepodvižnoju, tak že nepreravno, kak dviženie vremeni, dvigaloš čelovečstvo. Sлагались, razlagались različnye grupy ljudskih sceplenij; podgotovljались pričiny obrazovaniija i razloženiija gosudarstv, peremeščeniij narodov.

**Figure 8** [mode=ru\_cz,hyphenate=cs] *Czech phonetic transcription for contemporary Russian.*

Увьеть мужъ мужа, то мъститъ брату брата, или сынови отца, лубо отцо сына, или брату чаду, лубо сестрину сынови; аще не будеть кто мъститъ, то 40 гривенъ за голову; аще будеть русинъ, лубо гридинъ, лубо купчина, лубо лаветникъ, лубо мечникъ, аще изьгон будеть, лубо словенинъ, то 40 гривенъ положить за нь.

Uвьеть mužъ muža, to mьstitь bratu brata, ili synovi otca, ljubo oteču syna, ili bratu čadu, ljubo sestrinu synovi; ašče ne budetъ kto mьstitъ, to 40 grivenъ za golovu; ašče budetъ rusinъ, ljubo gridinъ, ljubo kupčina, ljubo lъbetnik, ljubo mečnikъ, ašče izsgoi budetъ, ljubo sloveninъ, to 40 grivenъ položitъ za нь.

**Figure 9** [mode=ocs\_cz,hyphenate=cs] *Czech phonetic transcription for Old Slavonic (superset of the corresponding Russian transcription).*

#### SERBIAN

The tables for converting Serbian text between Cyrillic and Latin alphabets are `sr_tolt` and `sr_tocy`.

Српски језик је један од словенских језика из породице индоевропских језика. Први писани споменици у српској редакцији старословенског језика потичу из XI и XII века. Српски језик је стандардни језик у службеној употреби у Србији, Босни и Херцеговини и Црној Гори, а у употреби је и у другим земљама где живе Срби, међу осталима и у Хрватској.

Srpski jezik je jedan od slovenskih jezika iz porodice indoevropskih jezika. Prvi pisani spomenici u srpskoj redakciji staroslovenskog jezika potiču iz XI i XII veka. Srpski jezik je standardni jezik u službenoj upotrebi u Srbiji, Bosni i Hercegovini i Crnoj Gori, a u upotrebi je i u drugim zemljama gde žive Srbi, među ostalima i u Hrvatskoj.

**Figure 10** [mode=sr\_tolt,hyphenate=hr] *Transliteration Кириллица → латиница.*

Srpski jezik je jedan od slovenskih jezika iz porodice indoevropskih jezika. Prvi pisani spomenici u srpskoj redakciji staroslovenskog jezika potiču iz XI i XII veka.

Srpski jezik je standardni jezik u službenoj upotrebi u Srbiji, Bosni i Hercegovini i Crnoj Gori, a u upotrebi je i u drugim zemljama gde žive Srbi, među ostalima i u Hrvatskoj.

Српски језик је један од словенских језика из породице индоевропских језика. Први писани споменици у српској редакцији старословенског језика потичу из XI и XIII века.

Српски језик је стандардни језик у службеној употреби у Србији, Босни и Херцеговини и Црној Гори, а у употреби је и у другим земљама где живе Срби, међу осталима и у Хрватској.

**Figure 11** [mode=sr\_tocy,hyphenate=sr] *Transliteration latinica → ćirilica.*

#### BULGARIAN

Българският език е индоевропейски език от групата на южнославянските езици. Той е официалният език на Република България и един от 23-те официални езика на Европейския съюз.

Bългарskijat ezik e indoevropski ezik ot grupata na južnoslavjanskite ezici. Toj e oficialnijat ezik na Republika Bălgarija i edin ot 23-te oficialni ezika na Evropejskija sąjuz.

**Figure 12** [mode=bg\_de,hyphenate=cs] *German scientific transliteration for Bulgarian (based on old iso 9 standard).*

#### LEGACY NATIONAL TRANSCRIPTIONS

At the moment there are tables for “old school” transcription into three languages: English (via `ru_transcript_en`), German (`ru_transcript_de`) and Czech (`ocs_cz`). At least the German one is almost unreadable if used with strings longer than two words. As we have the bijective ISO 9 mapping at hand there should be no reason at all to use any of them.

Прошло семь лет после 12-го года. Взволнованное историческое море Европы улеглось в свои берега. Оно казалось за- тихшим; но таинственные силы, двигающие человечество (таинственные потому, что законы, определяющие их движение, неизвестны нам), продолжали свое действие. Несмотря на то, что поверхность исторического моря казалась неподвижною, так же непрерывно, как движение времени, двигалось человечество. Слагались, разлагались различные группы людских сцеплений; подготавливались причины образования и разложения государств, перемещений народов.

Proshlo sem let posle 12-go goda. Vzvolo- vannoye istoricheskoye more Yevropy uleg- los v svoi berega. Ono kazalos zatikhshim; no tainstvennyye sily, dvigayushshiye chelovechestvo (tainstvennyye potomu, chto zakony, opredelyayushshiye ikh dvizheniye, neizvestny nam), prodolzhal svoye deystviye. Nesmotrya na to, chto poverkhnost istoricheskogo morya kazalas nepodvizhnoyu, tak zhe nepreryvno, kak dvizheniye vremeni, dvigalos chelovechestvo. Slagalis, razlagalis razlichnyye grupy lyudskikh stsepleni; podgotovlyalis prichiny obrazovaniya i razlozheniya gosudarstv, peremeshsheny nar- odov.

**Figure 13** [mode=ru\_transcript\_en,hyphenate=en] *English transcrip- tion for contemporary Russian.*

Прошло семь лет после 12-го года. Взволнованное историческое море Европы улеглось в свои берега. Оно казалось за- тихшим; но таинственные силы, двигающие человечество (таинственные потому, что законы, определяющие их движение, неизвестны нам), продолжали свое действие. Несмотря на то, что поверхность исторического моря казалась неподвижною, так же непрерывно, как движение времени, двигалось человечество. Слагались, разлагались различные груп- пы людских сцеплений; подготавливались причины образования и разложения госу- дарств, перемещений народов.

**Figure 14** [mode=ru\_transcript\_de,hyphenate=deo] *German transcrip- tion for contemporary Russian.*<sup>9</sup>

<sup>9</sup> Following Drosdowski/Müller/Scholze-Stubenrecht/Wermke (1952) p. 82; all the canonical rules are implemented save one: *-ezo* and *-ozo* should resolve to *-ewo* and *-owo* respectively iff genitive endings. As this is a grammatical rather than graphetical criterion writing a substitution algorithm would amount to do natural language parsing. To make things worse this rule is phonetically confused as it would not take care of other contexts where *z* in those patterns is articulated as */v/* like for instance in *сегодня* (which is a historical genitive, though ...). So even if this could be implemented it would not be advisable to use such a rule.





## REFERENCES

- 1 Birnbaum, H. and Schaeken, J. (1999). *Altkirchenslavische Studien*, volume 2. München.
- 2 Bornemann and Risch (1978). *Griechische Grammatik*. 2. edition Frankfurt am Main.
- 3 Bringhurst, R (2008). *The Elements of Typographic Style*. 3.2 edition Point Roberts WA, Vancouver.
- 4 DIN Deutsches Institut für Normung e. V., editor (2001). *Bibliotheks und Dokumentationswesen*. Berlin/Wien/Zürich.
- 5 Drosdowski, Müller, Schulze-Stubenrecht and Wermke, editors (1991). *DU-DEN Rechtschreibung der deutschen Sprache*. 20. edition Mannheim et al.
- 6 Kirschbaum (2001). *Grammatik der russischen Sprache*. Berlin.
- 7 ISO International Organization for Standardization, editor (1995). *Information and documentation – Transliteration of Cyrillic characters into Latin characters – Slavic and non-Slavic languages* 2. edition..