The

TRANSLITERATOR

for ConTEXt

MANUAL

CONTENTS

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          = ru_old ru ru_transcript_de ru_transcript_de_exp
                    ru_transcript_en all sr_tolt sr_tocy bg_de iso9_ocs ocs
                    ocs_gla ru_cz ocs_cz gr gr_n
    hyphenate     = cs sk hr
    hinting       = yes no
    sr_exceptions = yes no
```

```
\transliterate [..,.=.,..]
                      OPTIONAL
*   inherits from \setuptransliterator
```

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:

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

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

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.

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 Epicuros' sayings like this:

\transliterate[mode=gr]{κακὸν ἀνάγκη, ἀλλ' οὐδεμία ἀνάγκη ζῆν

3

μετὰ ἀνάγκης}

which yields "kakon anagkē, all' udemia anagkē zēn meta anagkēs" in the pdf output. Alternatively there is an environment, \starttransliterate[#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

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 TeX 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 "pronounceable" 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*.[1] 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

---

[1]   ISO (1995).

on the original language. Although fifteen years old it has not yet made its way into scholarly publications at large so it might not immediately look familiar.[2] The diacritics are not identical to the "scientific" transliteration used in Slavic studies but as long as your editor does not enforce its traditional method you should always prefer ISO 9 (`[mode=ru]`, `[mode=ru_old]`, `[mode=all]`).

But ISO 9, too, has its shortcomings. It has no definitions for historical forms of the cyrillic script like pre-XVIII-century Russian and Old (Church) Slavonic while those are covered by the scholarly transliterations. To amend the situation the Transliterator provides an extension to ISO 9 for Old Slavonic containing the glyphs ꙗ, ѕ, ѫ, ѭ, ѹ, ꙑ, ꙋ, ѡ, ѥ, ӻ, ѿ, ѱ and ѧ taken from the scientific transliteration (`[mode=iso9_ocs]`). If you prefer more coherency you might want to use pure "scientific" transliteration (`[mode=ocs]`). This method is complemented by `[mode=ocs_gla]`, the only option the Transliterator offers for the Glagolitic alphabet; they can be used consistently along each other as they were taken from the same book.[3]

As far as I know there is no standardized transliteration for Greek so I had to resort to the one that is used in scholarly literature. Its main drawback is that it has no representation for diacritics apart from (rough) breathing, but it respects specific rules for diphthongs and vowels in initial positions (`[mode=gr]`). There is one alternative mode for those who prefer their γ phonetically resolved to $/n/$ before velars (γ, $\varkappa$, χ and ξ; `[mode=gr_n]`).

Concerning the hyphenation within transliterated passages the default is set to to `[hyphenate=cs]` (Czech) which produces reasonable results when using `all`, `iso9_ocs` or `ru_cz`. For stuff like the English and German transcription use their respective native hyphenation.[4] However, as there is no hyphenation pattern I know of that closely resembles the transliteration of Greek you might have to resort to putting `\discretionary` hyphens when line breaking does not satisfy.

---

[2] A hasty glance at the latest issues of around 20 journals in a local library revealed that 2 of them actually are using ISO 9, these are *Przegląd wschodni* as of Nr. X, 3 (2008) and *Kwartalnik historyczny* as of CXVI, 3 (2009); the latter even contains a table on p. 218 showing a subset of the ISO 9 transliteration rules.

[3] Birnbaum/Schaeken (1999) p. 77 [http://www.schaeken.nl/lu/research/online/publications/akslstud/index.htm].

[4] You'll have to specify this through `\setuptransliterator` or locally because the default hyphenation is *not* the same as your documents'.

The Transliterator as a whole is nothing more than a bunch of dictionaries containing substitution rules for tokens that may occur in the text. These tokens may be single characters or strings of more than one character. As 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	ConTEXt
<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 ConTEXt 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 Transliterator should have little impact on document processing time at large, with the exception of the German transcription mode, perhaps.[5] 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.[6]

---

[5] 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.

[6] 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`).[7]

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 dovoľno krasivaâ ressornaâ neboľšaâ brička, v kakoj ezdât holostâki: otstavnye podpolkovniki, štabs-kapitany, pomešiki, imeûŝie okolo sotni duš krestʹân, — 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; neľzâ skazatʹ, čtoby star, odnako ž 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 полтіна, а не сотцкого, ино четырѣ гривны, а намъ вірь не таити в Новѣгородѣ; а о убіствѣ вірь нѣтъ. А что волости, честны король, новгородцкіе, ино тебѣ не держати своими мужи, а держати мужми новогородцкими. А что пошлина в Торжку и на Волоцѣ, тівунъ свои держати на своеи чясті, а Новугороду на своеи чясти посадника держаті. А се волости новогородцкіе: Волокъ со всѣми волостьми, Торжокъ, Бѣжіці, Городець Палець, Шіпинъ, Мелеця, Егна, Заволочье, Тиръ, Пермь, Печера, Югра, Вологда с волостьмі.

A svedetsâ vìra, ubʹût″ sotckogo v selě, ino tebě vzâti poltìna, a ne sotckogo, ino četyrě grivny, a nam″ vìr″ ne taiti v Nověgorodě; a o ubìstvě vìr″ nět″. A čto volosti, čestny koroľ, novgorodckìe, ino tebě ne deržati svoimi muži, a deržati mužmi novogorodckimi. A čto pošlina v Toržku i na Volocě, tìvun″ svoi deržati na svoei čâstì, a Novugorodu na svoei čâsti posadnika deržatì. A se volosti novogorodckìe: Volok″ so vsěmi volostmi, Toržok″, Běžìcì, Gorodecʹ Palecʹ, Šìpin″, Melecâ, Egna, Zavoločʹe, Tir″, Permʹ, Pečera, Ûgra, Vologda s volostmì.

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

---

[7] Unfortunately there are not yet any language files for some of them so please excuse the inadequate hyphenation in these cases.

Беларуская мова, мова беларусаў, увах-
одзіць у сям'ю індаеўрапейскіх моў, яе
славянскай групы і ўсходнеславянскіх мо-
ваў падгрупы, на якой размаўляюць у Бе-
ларусі і па ўсім свеце, галоўным чынам
у Расіі, Украіне, Польшчы. Б. м. падзяляе
шмат граматычных і лексічных уласцівас-
цяў з іншымі ўсходнеславянскімі мовамі
(гл. таксама: Іншыя назвы беларускай мо-
вы і Узаемныя ўплывы усходнеславянскіх
моваў).

Belaruskaâ mova, mova belarusaû, uvahodz-
ìc′ u sâmû ìndaeŭrapejskìh moŭ, âe slavâns-
kaj grupy ì ŭshodneslavânskìh movaŭ pad-
grupy, na âkoj razmaŭlâûc′ u Belarusì ì pa
ŭsìm svece, galoŭnym čynam u Rasìì, Ukra-
ìne, Pol′ŝčy. B. m. padzâlâe šmat gramatyč-
nyh ì leksìčnyh ulascìvascâû z ìnšymì ŭshod-
neslavânskìmì movamì (gl. taksama: Ìnšyâ
nazvy belaruskaj movy ì Uzaemnyâ ŭplyvy
ushodneslavânskìh movaŭ).

Figure 3 `[mode=all,hyphenate=cs]` *The complete cyrillic mapping from ISO 9; transliterating Belarusian.*

Украї́нська мова (застарілі назви – руська
мова, проста мова [...]) – слов'янська
мова, державна в Україні та одна з трьох
«офіційних мов на рівних засадах» у не-
визнаній 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.

Ukraíns′ka mova (zastarìlì nazvi – rus′ka mo-
va, prosta mova [...]) – slov'âns′ka mova, der-
žavna v Ukraïnì ta odna z tr′oh «ofìcìjnih
mov na rìvnih zasadah» u neviznanìj Prid-
nìstrovs′kìj Moldavs′kìj Respublìcì. Za rìzni-
mi ocìnkami zagalom u svìtì ukraïns′koû mo-
voû govorit′ vìd 41 mln. do 45 mln. osìb, vo-
na vhodit′ do tret′ogo desâtka najpoširenìših
mov svìtu.

Figure 4 `[mode=all,hyphenate=cs]` *The complete cyrillic mapping from ISO 9; transliterating Ukrainian.*

Српски језик је један од словенских је-
зика из породице индоевропских језика.
Први писани споменици у српској редак-
цији старословенског језика потичу из XI
и XII века. Српски језик је стандардни је-
зик у службеној употреби у Србији, Босни
и Херцеговини и Црној Гори, а у употре-
би је и у другим земљама гдје живе Срби,
међу осталима и у Хрватској.

Srpski jezik je jedan od slovenskih jezika
iz porodice indoevropskih jezika. Prvi pisa-
ni spomenici u srpskoj redakcìï starosloven-
skog jezika potìču iz XI i XII veka. Srpski
jezik je standardni jezik u službenoj upotre-
bi u Srbìjì, Bosni i Hercegovini i Crnoj Gori,
a u upotrebi je i u drugim zeml̂ama gdje žive
Srbi, među ostalima i u Hrvatskoj.

Figure 5 `[mode=all,hyphenate=cs]` *The complete cyrillic mapping from ISO 9; transliterating Serbian.*

9

Что са дѣютъ по въремьнемь : то ѿидето по вѣрь-
мьнемь : приказано боудѣте добрымъ людѣмъ :
а любо грамотою оутвѣрдать : како то боудѣте
всемъ вѣдомъ : или кто посль живыи ѿстанетъ
са : того лѣтꙷ коли алъбрахтъ : влⷣка ризкии
оумⷯрлъ : оуздоумалъ кнѧзѣ смольнескыи мьсти-
славъ : дⷣвⷣъ снⷮъ : прислалъ въ ригоу своего лоучꙋ-
шего попа : ерьмея : и съ нимь оумⷩа моужа пан-
телья : исвоего горда смольнеска : та два была по-
слъмь оу ризѣ : из ригы ехали на гоукыи берьго :
тамо твердити миръ :

Čto sę dějětě po věrem′nem′ : to otideto po
věr′m′nem′ : prikazano buděte dobrym″ lû-
děm″ : a lûbo gramotoû utvĕrdęt′ : kako to
buděte vsem″ vědom″ : ili kto posl′ živyi ôs-
tanĕt′ sę : togo lĕt̃′koli al″braht″ : vld̃ka rizkii
um″rl″ : uzdumal″ knęžě smol′neskyi : m′sti-
slav″ : dvd̃v″ sñ″ : prislal″ v″ rigu svojego
lučšego popa : jer′meâ : i s″ nim′ um′na muža
pantel′â : isvojego gorda smol′neska : ta dva
byla posl″m′ u rizě : iz rigy jehali na gočkyi
ber′go : tamo tverditi mir″ :

Figure 6 [mode=iso9_ocs,hyphenate=cs] *Transliteration rules according 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 научнемъ повѣсть сию. По потопѣ . первие снве
Ноеви . раздѣлиша землю . Симъ . Хамъ . Афетъ .
и таса въстокъ . Симови Персида . Ватрь . тоже и до
Индикиꙗ в долготу и в широу [и до Нирокоурия]
ꙗкоже реци ѿ въстока и до полоуденыꙗ . и Суриꙗ
. и Индиꙗ по Ефратъ рѣкоу . Вавилонъ . Кордуна
. Асуране . Мисопотамира . Аравиꙗ . старѣишаꙗ .
Елмансъ . Инди . Равиꙗ . на всꙗ Д.

Se načnemъ pověstь siju. Po potopě . pervie
snve Noevi . razděliša zemlju . Simъ . Хamъ
. Afetъ . i jasę vъstokъ . Simovi Persida .
Vatrъ . tože i do Indikija v dolgotu i v širinu
[i do Nirokuria] jakože rešti ot vъstoka i do
poludenьja . i Surija . i India po Efratъ rěku
. Vavilonъ . Korduna . Asurętne . Misopota-
mira . Aravija . starěišaja . Elmaisъ . Indi .
Ravija . na vsę D.

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

---

[8] 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; подготовлялись причины образования и разложения государств, перемещений народов.

Prošlo semь let posle 12-go goda. Vzvolnovannoe istoričeskoe more Evropy uleglosь v svoi berega. Ono kazalosь zatichšim; no tainstvennye sily, dvigajuščie čelovečestvo (tainstvennye potomu, čto zakony, opredeljajuščie ich dviženie, neizvestny nam), prodolžali svoe dejstvie. Nesmotrja na to, čto poverchnostь istoričeskogo morja kazalasь nepodvižnoju, tak že nepreryvno, kak dviženie vremeni, dvigalosь čelovečestvo. Slagalisь, razlagalisь različnye gruppy ljudskich sceplenij; podgotovljalisь pričiny obrazovanija i razloženija gosudarstv, peremeščenij narodov.

Figure 8 `[mode=ru_cz,hyphenate=cs]` *Czech phonetic transcription for contemporary Russian.*

Убьеть мужь мужа, то мьстить брату брата, или сынови отца, любо отцю сына, или братучаду, любо сестрину сынови; аще не будеть кто мьстя, то 40 гривень za голову; аще будеть русинъ, любо гридинъ, любо купчина, любо iабетник, любо мечникъ, аще изъгои будеть, любо словенинъ, то 40 гривень положити za нь.

Ubьetь mužь muža, to mьstitь bratu brata, ili synovi otca, ljubo otcju syna, ili bratučadu, ljubo sestrinu synovi; ašče ne budetь kto mьstię, to 40 grivenъ za golovu; ašče budetь rusinъ, ljubo gridinъ, ljubo kupčina, ljubo iębetnik, ljubo mečnikъ, ašče izъgoi budetь, ljubo sloveninъ, to 40 grivenъ položiti za nь.

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 ћирилица → 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 ХИ и ХИИИ 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.

Bălgarskijat ezik e indoevropejski ezik ot grupata na južnoslavjanskite ezici. Toj e oficialnijat ezik na Republika Bălgarija i edin ot 23-te oficialni ezika na Evropejskija săjuz.

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 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; подготовлялись причины образования и разложения государств, перемещений народов.

Proshlo sem let posle 12-go goda. Vzvolnovannoye istoricheskoye more Yevropy uleglos v svoi berega. Ono kazalos zatikhshim; no tainstvennyye sily, dvigayushshiye chelovechestvo (tainstvennyye potomu, chto zakony, opredelyayushshiye ikh dvizheniye, neizvestny nam), prodolzhali svoye deystviye. Nesmotrya na to, chto poverkhnost istoricheskogo morya kazalas nepodvizhnoyu, tak zhe nepreryvno, kak dvizheniye vremeni, dvigalos chelovechestvo. Slagalis, razlagalis razlichnyye gruppy lyudskikh stsepleniy; podgotovlyalis prichiny obrazovaniya i razlozheniya gosudarstv, peremeshsheniy narodov.

Figure 13 `[mode=ru_transcript_en,hyphenate=en]` *English transcription for contemporary Russian.*

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; подготовлялись причины образования и разложения государств, перемещений народов.

Proschlo sem let posle 12-go goda. Wswolnowannoje istoritscheskoje more Jewropy uleglos w swoi berega. Ono kasalos satichschim; no tainstwennyje sily, dwigajuschtschije tschelowetschestwo (tainstwennyje potomu, tschto sakony, opredeljajuschtschije ich dwischenije, neiswestny nam), prodolschali swoje deistwije. Nesmotrja na to, tschto powerchnost istoritscheskogo morja kasalas nepodwischnoju, tak sche neprerywno, kak dwischenije wremeni, dwigalos tschelowetschestwo. Slagalis, raslagalis raslitschnyje gruppy ljudskich schepleni; podgotowljalis pritschiny obrasowanija i rasloschenija gossudarstw, peremeschtscheni narodow.

Figure 14 `[mode=ru_transcript_de,hyphenate=deo]` *German transcription for contemporary Russian.*[9]

GLAGOLITIC

[vl] adъiko čkъ si xoštetъ raz[orit] i zakonъ manastъirъski: [iže] tъi ustavi Reče že igum[ъ] [nъ] kako xoštetъ razoriti zak[onъ] [.] [a] estъ· črъvi i [m] [..........] [..] mъ ne možetъ sego trъpĕt[i] [da] lubo sego imĕi sъdĕ a mъi o [ti]demъ: lubo sego pusti: da ot [id]etъ iže estъ prišъlъ: s[e]

Figure 15 `[mode=ocs_gla,hyphenate=cs]` *"Scientific" transliteration for Old Slavonic written in the Glagolitic alphabet as used in Birnbaum/Schaeken (1999).*

GREEK

The Transliterator offers two modes for handling Greek: `gr` and `gr_n`. They differ only on one aspect. `gr` transliterates the canonical Greek alphabet as well as the special glyphs Digamma, Quoppa and Sampi. `gr_n` behaves exactly the same way except that nasalization is observed such that γ+[γ|κ] yields n+[g|k].

---

[9] Following Drosdowski/Müller/Scholze-Stubenrecht/Wermke (1952) p. 82; all the canonical rules are implemented save one: *-его* and *-ого* 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 *г* 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.

οἴνῳ δὲ κάρτα προσκέαται, καί σφι οὐκ ἐμέσαι ἔξεστι, οὐδὶ οὐρῆσαι ἀντίον ἄλλου. ταῦτα μέν νυν οὕτω φυλάσσεται, μεθυσκόμενοι δὲ ἐώθασι βουλεύεσθαι τὰ σπουδαιέστατα τῶν πρηγμάτων· τὸ δ᾽ ἂν ἅδῃ σφι βουλευομένοισι, τοῦτο τῇ ὑστεραίῃ νήφουσι προτιθεῖ ὁ στέγαρχος, ἐν τοῦ ἂν ἐόντες βουλεύωνται, καὶ ἢν μὲν ἅδῃ καὶ νήφουσι, χρέωνται αὐτῷ, ἢν δὲμὴ ἅδῃ, μετιεῖσι. τὰ δ᾽ ἂν νήφοντες προβουλεύσωνται, μεθυσκόμενοι ἐπιδιαγινώσκουσι.

oinō de karta proskeatai, kai sphi uk emesai exesti, uki urēsai antion allu. tauta men nyn hutō phylassetai, methyskomenoi de eōthasi buleuesthai ta spudaiestata tōn prēgmatōn: to d' an hadē sphi buleuomenoisi, tuto tē hysteraiē nēphusi protithei ho stegarchos, en tu an eontes buleuōntai, kai ēn men hadē kai nēphusi, chreōntai autō, ēn demē hadē, metieisi. ta d' an nēphontes probuleusōntai, methyskomenoi epidiaginōskusi.

Figure 16   `[mode=gr,hyphenate=de]`   *Transliteration for Greek – standard.*

ταῦτα καὶ νεωτέρῳ καὶ πρεσβυτέρῳ ὅτῳ ἂν ἐντυγχάνω ποιήσω, καὶ ξένῳ καὶ ἀστῷ, μᾶλλον δὲ τοῖς ἀστοῖς, ὅσῳ μου ἐγγυτέρω ἐστὲ γένει.

tauta kai neōterō kai presbyterō hotō an entynchanō poiēsō, kai xenō kai astō, mallon de tois astois, hosō mu engyterō este genei.

Figure 17   `[mode=gr_n,hyphenate=de]`   *Transliteration for Greek – alternative respecting nasalization.*

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). *DUDEN 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..