教会スラヴ語のテキストデータ作成方法

Methods for Creating Church Slavonic Textual Data

Методы составления текстовых данных церковнославянского языка

北大言語学サークル Huling (北海道大学 非公認団体)

HP: huling.org e-mail: linguistic.society.of.hu@gmail.com

文責:宮西大

2023年10月30日

※本文書に記載した URL はすべて 2023 年 10 月 30 日現在のものである.

1. 概要

本文書は、教会スラヴ語北大言語学サークル Huling が HP にて公開している「教会スラヴ語キリル文字変換器」の利用手引きである。また付記にて、Huling 以外の個人・団体が作成した教会スラヴ語関連のツールの通覧もまとめている。

1.1 教会スラヴ語について

本文書では、古教会スラヴ語 Old Church Slavonic と教会スラヴ語 Church Slavonic という2つの言語のテキストデータ化について取り扱っている。これらの言語は主に羊皮紙の写本に書かれ、今日まで残されているスラヴの古い文章語である。ところで「2つの言語」と述べはしたものの、両者の間に決定的な差異があるわけではない。この両言語の区分は学術上定義づけられた便宜的なものであり、前者は1100年以前の主にブルガリア周辺で書かれた教会典礼語を、後者は1101年以降の方言的差異が顕著化したスラヴ各地の典礼語を指す。スラヴ祖語が言語としての統一性を最終的に失ったのがおおよそ12世紀初頭だと見積もられているため、人為的にこのような線引きがなされている。

しかしテキストデータ化という観点で見れば、両言語を区別する必要性はほとんどないと思われる. 文字形状のヴァリアントや書体は時代ごとに異なり、文献学上無視できない問題だと思われるが、本文書のキーワードであるテキストデータとはつまりコンピュータ上で扱う文字コードの情報のみであるので、ここでは文字の特徴に関する立ち入った議論は行なわない。

したがって、以下本文書では特に断りを入れない限り、古教会スラヴ語と教会スラヴ語をひとまとめに単に「教会スラヴ語」と呼ぶことにする。

1.2 文字変換器について

文字変換器とは、非 ASCII 文字の取得を目的とした Huling の Web ツールである. 想定される利用目的は以下の 2 通りである.

- ① ASCII 文字をダイアクリティカル・マーク付きラテン文字や非ラテン文字(ギリシア文字やキリル文字, ブラーフミー系文字, ハングルなど)に変換し取得する.
 - ② 非ラテン文字をラテン文字(ダイアクリティカル・マーク付き含む)に翻字する.

文字変換器の主たる目的は前者であり、後者は変換器の仕組みを利用して生まれた副次的なものである. 翻字規則には様々な流儀があるが、ここで採用しているのはそのうちのひとつのみであり、ユーザが自由に選ぶことはできない. 非ラテン文字からラテン文字への翻字は、あくまで補助的な機能であると理解していただきたい.

文字変換器は2つのテキストボックスから成り、基本的に一方がラテン文字、もう一方が非ラテン文字の入出力に使われる. いずれかのボックスを選択するとそこに書かれてある文字列が入力となり、何らかの翻字規則を適用して変換した後に隣のボックスに文字列が出力される.

Huling は教会スラヴ語の入力を念頭に置いた文字変換器として「教会スラヴ語キリル文字」と「グラゴル文字」の2つを公開しているが、本文書では機能が比較的充実しているキリル文字版のみについて解説する.

1.3 教会スラヴ語のテキストデータ化・組版処理

このような事情の下で、どのようにして教会スラヴ語をコンピュータ上で扱うか、様々な方策が練られてきた。その際に問題となるのが、どの程度まで実物の写本を再現するかということである。

最も簡易で汎用性の高い手段は、キリル文字をASCII 文字に翻字して記述するものである。このとき、字体などの特徴はすべて捨象され、写本に書かれている文字の配列のみが情報として残される。言語研究の資料としてはこの程度でも十分利用可能であり、例えば1986年から2017年までヘルシンキ大学で作成されたCorpus

^{1 15}世紀頃までの古い時代の文献やクロアチア沿岸部で19世紀初頭までに作成された文献に関しては、グラゴル文字が使用されている例が多数存在する.しかし、そのような場合においても特に学術研究ではグラゴル文字はすべてキリル文字に翻字されるのが慣例となっており、実用上グラゴル文字が必要とされることはあまりない.

Cyrillo-Methodianum Helsingiense (CCMH)² はこの方法によっている。あるいは旧ソ連圏であれば拡張 ASCII KOI-8³に収録されているキリル文字を用いることも可能である。しかしいずれにしても,ASCII を土台にした文字集合では教会スラヴ語の表記に必要な文字を大文字・小文字や異体字の区別を含めてすべて符号化することなど現実的に不可能である。

そこで次なる手段として開発されたのが組版処理技術である。 教会スラヴ語の主要な TEX パッケージとしては 2000 年公開の HipTeX が挙げられる。 また 1990 年代には SlavTeX というパッケージも公開されていたようである。 教会スラヴ語 TeX に関しては安田功氏による以下のサイトに詳しい:

しかし 2000 年代後半以降 Unicode が急速に普及し、キリル文字の収録数も充実してくると、教会スラヴ語のテキストはある程度そのままキリル文字で記述できるようになった。教会スラヴ語のテキストデータ化に必要な文字の大半は、2008 年の Unicode 5.1.0 で既に収録済みである。しかし実際の運用上これが活用されているかと言えば、必ずしもそうではない。例えば U+A657 ta の文字の代わりに U+044F g を使い、フォントで ta のように見せかけている Web サイトも存在する。 Huling の文字変換器は、このような状況を僅かながら変えていくことを期待して作成された。 Web 上で誰でも容易に ASCII 文字入力から古いキリル文字を取得できるのは、文字変換器の強みだと考えられる。

学術的・技術的なご指摘やご助言があれば、HPの当該ページ下部にあるコメント欄または題記のメールアドレスまでご連絡いただければまことに幸いである.

² https://korp.csc.fi/download/ccmh-src/www/index.html

³ Код Обмена Информацией, 8 бит の略

2. 文字変換器 利用の手引き

2.1 ASCII 入力早見表

文字変換器の主たる目的である、ASCII 文字から古いキリル文字を取得するための対応表を以下に示す。「ASCII」の列にスペースで区切られた2つの例が記されている場合は、その両方が入力可能となっている。

表 1: 基本字母の ASCII 入力

キリル文字	ASCII	キリル文字	ASCII	キリル文字	ASCII	キリル文字	ASCII
А	Α	۵	a	Оу	U	٥y	u
Б	В	Б	b	Ф	F	ф	f
В	V	В	V	Θ	Th	.0.	th
Γ	G	Г	g	X	X	X	x
Д	D	А	d	w	W	w	W
6	Е	E	e	ті	Sht	щ	sht
Ж	Zh	ж	zh	Ц	С	ц	С
S	Dz	s	dz	Ч	Ch	ч	ch
ζ	Z	3	z	Ш	Sh	ш	sh
3	Z∖	3	z\	Ъ	QQ	ጌ	qq
I	J	1	j	ы	Υ	ъ	у
ι	Јj	ι	jj	Ь	Q	k	q
И	I	И	i	ቴ	Eh	ቴ	eh
<u></u>	Gh	, t	gh	Ю	Ju	ю	ju
K	K	к	k	M	Ја	ra	ja
Л	L	Λ	1	Æ	Je	ıє	je
М	М	M	m	А	EN	A	eN @@
N	N	н	n	Ж	ON	ж	oN @
0	0	0	o	Ы	JeN	ЬA	jeN j@@
П	Р	п	р	₩	JoN	hX.	joN j@
ρ	R	ρ	r	ğ	Ks\	ă	ks\
\mathbf{e}	S	c	S	Ψ	Ps∖	ψ	ps\
Т	Т	т	t	V	Uh	v	uh

表 2: その他字母の ASCII 入力

キリル文字	ASCII	キリル文字	ASCII	キリル文字	ASCII
$oldsymbol{\mathcal{K}}$	D\	Д	d\	А	d\\
6	Ε\	€	e\		
\mathbf{s}	Dz\	г	dz∖	8	v\
Z	Dz\\	3	dz\\		
ħ	Gh∖	ħ	gh∖		
Л	L\	х	1\		
W	M\	W	m\		
И	N\	N	n\		
Θ	0.\	•	0.\	0	0\
\odot	0.\	Θ	0.\	**	0.\\\\
∞	0.\\	®	0.\\	c	s\
$\mathbf{\omega}$	00\	00	00\	Т	t\
Ф	0x\	Ф	ox\	m	t\\
\aleph	U\	8	u\	γ	u\\
Ö	W\t	æ	w\t		
Ф	W\	Ф	w\		
co	W\\	ယ	w\\		
©	W\\\	ස	w\\\		
μ	C\	р	c\		
\mathbf{G}	Κ\	ና	k\		
Ţ	Dzh\	Ų	dzh∖		
Ъ	Qq\	ъ	qq\	ጌ	qq\\
د	Jeh∖	由	jeh\	ቴ	eh\
Ю	Ju\	он	ju\		
A	EN∖	А	eN\ @@\		
Ы	JeN\ J@@\	IA	jeN\ j@@\		
X	AN\	*	aN\		
	YN\	个	yN\		

表 3: 記号・約物のASCII入力

記号•約物	ASCII	
63	^< ^>	気息記号
ઁ૾૽	^Q ^QQ	イェルの省略記号
52	^~ ^~~	語の省略記号
$\widehat{\circ}$	۸۸	語の省略記号
ô	^3	口蓋化の記号
⊽Ō	^- ^=	オーバーライン
ć ″	^/ ^//	アキュート・アクセント
` "	^\ ^\\	グレイヴ・アクセント
ö	۸٬٬٬	ダイアレシス
·: :·	.: :.	
*::	:: ::\	
∺ +	::. ::.\	
∴-	.:- ::- etc.	
R	[Mr]	"Symbol for Marks Chapter"
‡	[Kr]	八端十字架
;	;\	二重コンマ(対応フォント少)
* #	# #\	米印
¶	\ P	段落記号
« »	<< >>	引用符
Ť	+\	短剣符

表 4: 上付き結合文字の ASCII 入力

 結合文字	ASCII	結合文字	ASCII
ô	^a	\$	^f
់	^b	ô	^th
ឺ	^v	č	^x ^h
Ö	^g	ů Č	^W
ै	^d	"	^sht
ំ	^e	Ö	^c
ж О	^zh	ឺ	^ch
ô	^z		^sh
ů O	^i	৾	^qq
ô	^gh	<u>ы</u>	^y
к О	^k	ំ	^q
ô	^1	Q. A.	^eh
ő	^m	Ö	^ju
N O	^n	ta O	^ja
ဝိ	^o	ю О	^je
n O	^ p	â	^eN ^@@
်	^r	ő	^oN ^@
ំ	^s	ых О	^joN ^j@
៊	^t	č	^uh
ំ	^u		

2.2 文字対応の詳細

文字変換器で取得できるキリル文字の具体的なUnicode の符号位置および対応するASCII 文字入力, ダイアクリティカル・マーク付きラテン文字を以下に示す.

表 5: 文字対応の詳細

符号位置	文字	ASCII	翻字	
U+0410	А	А	Α	
U+0430	a	а	a	
U+0411	Б	В	В	
U+0431	Б	b	b	
U+0412	В	V	V	
U+0432	В	V	V	
U+1C80	в	v\	V_1	丸いm
U+0413	Γ	G	G	
U+0433	Г	g	g	
U+0414	Д	D	D	
U+0434	А	d	d	
U+1C81	А	d\\	d ₂	ひげの長い _A
U+A662	Д	D\	D_1	軟らかい _A
U+A663	Д	d\	d ₁	事人 のガ ヤ、み
U+0415	6	E	Е	
U+0435	E	e	e	
U+0404	6	E∖	E ₁	F1. \ .
U+0454	ϵ	e\	e ₁	広い・・
U+0416	Ж	Zh	Ž	

符号位置	文字	ASCII	翻字	名称•備考
U+0436	ж	zh	ž	
U+0405	S	Dz	3	
U+0455	S	dz	3	
U+A642	ξ	Dz\\	32	
U+A643	3	dz\\	32	
U+A644	S	Dz\	31	
U+A645	s	dz∖	31	
U+0417	3	Z∖	Z_1	
U+0437	3	z\	Z ₁	
U+A640	ζ	Z	Z	
U+A641	3	z	Z	
U+0418	И	I	I	
U+0438	И	i	i	
U+0406	I	J	Ì	
U+0456	1	j	ì	
U+A646	ι	Ͻj	l	
U+A647	ι	jj	ι	
U+A648	<u></u>	Gh	Ġ	
U+A649	*	gh	ģ	
U+040B	ħ	Gh∖	Ğ1	
U+045B	ħ	Gh\	ģ ₁	
U+041A	K	K	K	
U+043A	к	k	k	

符号位置	文字	ASCII	翻字	名称•備考
U+041B	Л	L	L	
U+043B	Λ	1	1	
U+A664	Λ	L\	L_1	軟らかいへ
U+A665	Т	1\	11	事人のカーV・A
U+041C	М	М	М	
U+043C	M	m	m	
U+A666	W	M\	M_1	軟らかいw
U+A667	W	m\	m_1	#90-977 W · M
U+041D	И	N	N	
U+043D	N	n	n	
U+04A4	И	N\	N_1	軟らかいw
U+04A5	И	n\	n_1	#X-9% V · H
U+041E	0	0	0	
U+043E	٥	O	0	
U+1C82	٥	0\	0 ₁	狭い 。
U+A668	0	0.\	-	
U+A669	•	0.\	-	
U+A66A	Θ	0.\\	-	
U+A66B	©	0.\\	-	
U+A66C	∞	0.\\\	-	
U+A66D	00	0.\\\	-	
U+A66E	**	0.\\\\	-	
U+A698	ω	00\	-	

符号位置	文字	ASCII	翻字	名称•備考
U+A699	w	00\	-	
U+A69A	Ф	00\\	-	
U+A69B	Ф	00//	-	
U+041F	П	Р	Р	
U+043F	П	р	р	
U+0420	Р	R	R	
U+0440	ρ	r	r	
U+0421	G	S	S	
U+0441	c	S	S	
U+1C83	c	s\	S ₁	広いc
U+0422	Т	Т	T	
U+0442	т	t	t	
U+1C84	Т	t\	t ₁	背の高い T
U+1C85	m	t\\	t ₂	三本足のT
U+041E U+0443	Оу	U	U	11-0479 0 11-0470) 计非性将
U+1C82 U+0443	ογ	u	u	U+0478 Oy, U+0479 oy は非推奨.
U+A64A	R	U\	U ₁	
U+A64B	8	u\	U ₁	
U+1C88	ጸ	u\\	U_2	
U+0423	У	Uh\	Ü₁	
U+0443	У	uh\	ü₁	
U+0424	Ф	F	F	

符号位置	文字	ASCII	翻字	名称•備考
U+0444	ф	f	f	
U+0425	X	x	Χ	
U+0445	X	x	Х	
U+0460	w	W	ω	
U+0461	w	W	ω	
U+A64C	ငာ	W\\	$\omega_{\text{\tiny 2}}$	
U+A64D	ယ	w\\	ω_{2}	
U+047C	@	W\\\	ω_{3}	
U+047D		w\\\	ω_3	
U+047E	Ö	W\t	ω^{t}	
U+047F	w	w\t	$\omega^{^{\mathrm{t}}}$	
U+0426	Ц	С	С	
U+0446	ц	С	С	
U+A660	Ц	C\	C_1	
U+A661	μ	c\	C ₁	
U+0427	Ч	Ch	Č	
U+0447	Ч	ch	č	
U+0480	\mathbf{G}	K\	K ₁	
U+0481	ና	k\	k ₁	
U+040F	Ţ	Dzh\	Dž ₁	
U+045F	Ų	Dzh\	dž₁	
U+0428	Ш	Sh	Š	
U+0448	Ш	sh	š	

符号位置	文字	ASCII	翻字	名称•備考
U+0429	ιjı	Sht	Št	III _T = Sh.t
U+0449	ψ	sht	št	шт = sh.t
U+042A	Ъ	QQ	Ъ	
U+044A	ጌ	qq	Ъ	
U+1C86	ጌ	qq\	Ъ ₁	背の高い ъ
U+A650	Ъ	Υ	Υ	
U+A651	ъ	у	У	
U+042B	Ы	Υ\	Y ₁	
U+044B	Ы	y\	y ₁	
U+042C	Ь	Q	Ь	
U+044C	Ь	q	Ь	
U+A64F	Ъ	qq\\	Ъ	
U+0462	ቴ	Eh	Ě	
U+0463	ቴ	eh	ě	
U+1C87	\$	eh\	ě₁	
U+A652	捁	Jeh\	Jě₁	
U+A653	由	Jeh\	jě₁	
U+042E	Ю	Ju	Ju	
U+044E	ю	ju	ju	
U+A654	Ю	Ju\	Ju₁	
U+A655	Ø	ju\	ju₁	
U+A656	M	Ја	Ja	
U+A657	ra	ja	ja	

符号位置	文字	ASCII	翻字	名称•備考
U+0464	16	Je	Je	
U+0465	Æ	je	je	
U+0466	A	EN	Ę	
U+0467	A	eN @@	ę	
U+A658	А	EN\	Ę1	
U+A659	Δ	eN\ @@\	ę₁	
U+046A	X	ON	Ó	
U+046B	Ж	oN @	Ó	
U+A65A	\mathbf{x}	AN\	Œ	
U+A65B	×	aN∖	œ	
U+A65E	Λ	YN\	ŶN	ルーマニア正教会で使用された文字
U+A65F	小	yN\	ŷn	ルーマーグ正教会で使用で40元又子
U+A65C	Ы	JeN\ J@@\	Ję₁	
U+A65D	Ы	jeN\ j@@\	ję₁	
U+046C	Ж	JoN J@	Jό	
U+046D	hX.	joN j@	jǫ	
U+046E	ğ	Ks\	Ks ₁	
U+046F	ž	ks\	ks ₁	
U+0470	Ψ	Ps\	Ps_1	
U+0471	Ψ	ps\	ps ₁	
U+0472	Θ	Th	Th	
U+0473	.0.	th	th	
U+0474	V	Uh	Ü	

符号位置	文字	ASCII	翻字	名称•備考	
U+0475	V	uh	ü		

3. 関連ツール・サイト

3.1 Slavonic Computing

リンク: sci.ponomar.net ponomar.net

サンクトペテルブルク神学アカデミーの Aleksandr Andreev 氏を中心としたグループが開発しているサイト. 教会スラヴ語の TeX やフリーフォント, 仮想キーボードなどを公開している.

3.2 Truetype フォント Igor'-in, Igor'-out

明治大学文学部の岩井憲幸教授および石川幹人教授が作成したフォント. 1998 年度から始まった研究「古代ロシア文語の萌芽期における特性の研究――『アルハンゲリスク福音書』を中心として」の一環として作られた. 作成当時はキリル文字の入力インターフェースも Unicode のような規格もまだ発展途上だったため, ひらがな・カタカナにキリル文字を割り付け, フォントによって見かけ上キリル文字を表示させるという代替手段を取っている. これについては岩井&石川(1999)『古代スラブ文字データベースの作成』『に詳しい. 表 6 は Igor'-in, Igor'-out フォントにおける仮名文字・ラテン文字とキリル文字との対応を示している. Igor'-in については岩井&石川(1999)を参照し, Igor'-out については恩田(2016)の pdf版から抽出されたテキストデータを基に筆者が分析した.

3.3 Microsoft Word

「記号と特殊文字」からキリル文字を入力することができる。しかし Word の特殊文字は、フォントが対応している場合のみ一覧表に表示されるようになっているため、あらかじめ目当ての文字のフォント対応状況を確認しておく必要がある。例えばフォントを Times New Roman に設定して U+A641 g を入力しようとしてもこれは文字欄には現れず、別途適当なフォントを用意しなければならない。このような方法は長大なテキストを入力するのにはまったく不向きである。

⁴ Github: typiconman, アカデミー公式 HP: Андреев Александр Андреевич

⁵ 明治大学学術成果リポジトリよりダウンロード可能

表 6: Unicode 符号位置, Igor'-in, Igor'-out の対応表

	Unicode	Igor'-in	Igor'-out		Unicode	Igor'-in	Igor'-out		Unicode	Igor'-in	Igor'-out
a	0430	あ	Α	M	043C	む	Υ	ц	0446	つ	Ξ
ä	0430 + 0308	あ	В	Ä	043C + 0308	ま	Z	ij,	0446 + 0308	っ	ヌ
Б	0431	<i>آ</i> گر،	С	И	043D	h	7	ч	0447	5	ネ
Ë	0431 + 0308	ば	D	й	043D + 0308	ね	7	ш	0448	L	/
В	0432	を	E	И	04A5	~" ~	1	Ъ	044A	ぬ	Λ
Ë	0432 + 0308	び	F	0	043E	お	ゥ	ъ	A651	な	٤
Г	0433	<	G	ö	043E + 0308	3	I	Ь	044C	ſこ	7
Ë	0433 + 0308	ぎ	Н	w	0461	お	ħ	ቴ	0463	め	^
Д	0434	ど	I	ü	0461 + 0308	ŧ	л	ю	044E	ゆ	朩
Ä	0434 + 0308	だ	J	w	047F	Ø	‡	ta	A657	ゃ	₹
ε	0435	え	K	п	043F	ιζι	7	ю	0465	え	Ę
ж	0436	じ	L	ρ	0440	る	τ	А	0467	わ	L
s	0455	ぢ	М	ö	0440 + 0308	6	٦	Ж	046B	ょ	X
3	0437	₫"	N	c	0441	す	Ħ	ЬA	0469	ゎ	ŧ
3	A641	づ	0	ë	0441 + 0308	ż	ý	Ѭ	046D	ţ	ヤ
ı	0456	ι\	Р	т	0442	٤	λ	ă	046F	き	ı
ι	A647	み	Q	Ϋ́	0442 + 0308	で	t	ψ	0471	Q_{c}	Ε
ï	0457	∇	R	٥y	1C82 + 0443	う	y	Ψ̈́	0471 + 0308	Œ	Ī
И	0438	ſ١	S	8	A64B	j	タ	v	0475	ゃ	IJ
ħ	045B	^	Т	ф	0444	ıZı	Ŧ	У	0443	ゆ	JV
к	043A	<	U	.0.	0473	τ	Ÿ	ና	0481	か	ν
Λ	043B	ħ	V	.ö.	0473 + 0308	ぼ	$\bar{\tau}$	s	A645	ざ	П
Ä	043B + 0308	Ŋ	W	Х	0445	ほ	 	Δ	2C51 (?) ⁶	ť"	7
Л	A665	が	Х	щ	0449	た	t	A	A659	ぞ	ン

⁶ U+2C51 はグラゴル文字 \blacktriangle (ě) のコードポイントである. 鼻母音 $\mathfrak e$ を表すキリル文字として Unicode 15.0 で収録されているのは現状 U+0467 \blacktriangle , U+A659 \blacktriangle およびそれぞれの大文字のみであり, \blacktriangle のような字形のキリル文字は収録されていない. Unicode Consortium に文字の追加を提案してみても良いかもしれない.

4. 参考文献

Andreev, A., Shardt, Y., & Simmons, N. (2015). Church Slavonic Typography in Unicode (1st ed.).

https://www.unicode.org/notes/tn41/tn41-1.pdf

岩井憲幸&石川幹人. (1999). 古代スラブ文字データベースの作成. 明治大学情報科学センター年報. 11 巻. p. 45-51. https://meiji.repo.nii.ac.jp/records/16934

恩田義徳. (2016). 古代教会スラブ語の分詞について 福音書テキストを対象に. 博士学位論文(東京外国語大学). http://repository.tufs.ac.jp/bitstream/10108/86462/1/dt-ko-0209.pdf 木村彰一. (2003). 古代教会スラブ語入門《新装版》. 白水社.