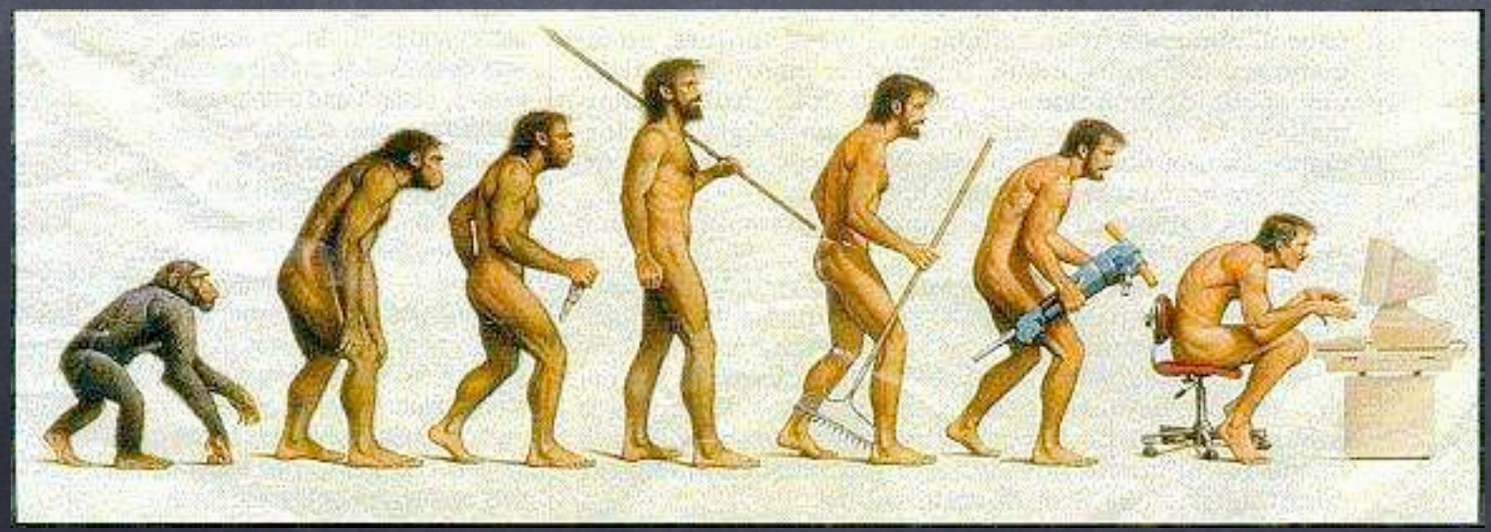
https://doi.org/10.20378/irb-49297

UNICODE 4.1 AND SLAVIC PHILOLOGY - PROBLEMS AND PERSPECTIVES -

Prof. Dr. Sebastian Kempgen
University of Bamberg, Germany

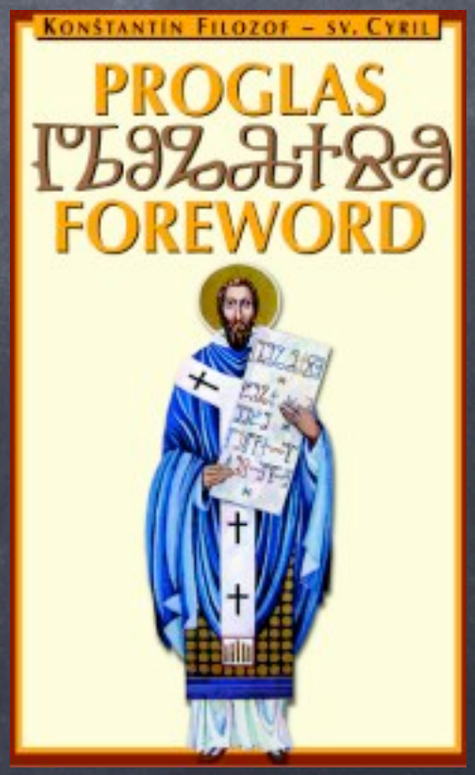
Overview

- Introduction
- Status Quo : What's Available
- Problems: Missing Pieces
- Perspectives: More Characters



Topics Covered

- Russsian Hist. Orthography
- Russian Phonetics
- Polish Hist. Orthography
- Sorbian Hist. Orthography
- Croatian Accents
- Bulgarian Phonetics
- Old Church Slavonic
- Transliteration of Glagolitic
- Balkan Philology



1. Introduction

- ASCII: 2^7 characters = 128
- 'Code Pages': 2^8 characters = 256
- Unicode: 2^16 characters = 65.536
- Versions: v. 1.0 1991, v. 4.1 2005
- Version 4.1: Glagolica ✓

Unicode Blocks

uni040(	Iocyril	Djecyr	Gjecyr	Ecyrill	Dzecyr	Icyrilli	Yicyril	Jecyri	Ljecyr	Njecyr	Tshecy	Kjecyr	uni040I	Ushort	Dzhecy
È	Ё	Ђ	Ѓ	Є	S	I	Ї	J	Љ	Њ	Ћ	Ќ	Й	Ў	Џ
Acyrill	Becyri	Vecyri	Gecyri	Decyri	Iecyril	Zhecyr	Zecyril	Iicyrill	Iishort	Kacyri	Elcyril	Emcyri	Encyril	Ocyrill	Pecyri
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Ercyril	Escyri	Tecyri	Ucyrill	Efcyril	Khacyr	Tsecyr	Checyr	Shacyr	Shchac	Hardsi	Yericy	Softsi	Erever	IUcyril	IAcyri
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
acyrill	becyri	vecyri	gecyri	decyri	iecyril	zhecyr	zecyril	iicyrill	iishort	kacyri	elcyril	emcyri	encyril	ocyrill	pecyri
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
ercyril	escyri	tecyril	ucyrill	efcyril	khacyr	tsecyr	checyr	shacyr	shchac	hardsi	yericy	softsi	erever	iucyril	iacyril
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
uni045(	iocyril	djecyr	gjecyr	ecyrill	dzecyr	icyrilli	yicyril	jecyril	ljecyri	njecyr	tshecy	kjecyr	uni045I	ushort	dzhecy
è	ё	ђ	ѓ	є	s	i	ї	j	љ	њ	ћ	ќ	й	ў	џ
Omega	omega	Yatcyr	yatcyr	Eiotifi	eiotifi	Yuslitt	yuslitt	Yuslitt	yuslitt	Yusbig	yusbig	Yusbigi	yusbigi	Ksicyr	ksicyri
Ѡ	ѡ	Ѣ	ѣ	Ѥ	ѥ	Ѧ	ѧ	Ѩ	ѩ	Ѫ	ѫ	Ѭ	ѭ	Ѯ	ѯ
Psicyr	psicyri	Fitacyr	fitacyr	Izhitsa	izhitsa	Izhitsa	izhitsa	Ukcyri	ukcyril	Omegar	omegar	Omegat	omegat	Otcyril	otcyril
Ѱ	ѱ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ	ѽ	Ѿ	ѿ
Koppac	koppac	thousa	titlocyr	palatal	dasiapr	psilipne	uni0487	uni0488	uni0489	lishort	iishort	Semiso	semiso	Ercyril	ercyril
Ҁ	ҁ	≠	◌҃	◌҆	⊢	⊣				Ҋ	ҋ	Ҍ	ҍ	Ҏ	ҏ

Unicode Blocks

- Basic Latin, Latin-1, Latin Extended-A, Latin Extended-B, Latin Extended Additional; Latin Extd C (5.0)
- Cyrillic, Cyrillic Extended (= Non-Sl.)
- Greek, Greek Extended (= Hist.)
- IPA, IPA Extensions + Supplement
- Spacing & Combining Diacritics

‘Private Area’

- 6.400 Slots
- Compatibility between fonts and documents not guaranteed
- Use at your own risk
- Already in use
- Use not coordinated between philologies

Unicode.org

- Submissions, Proposals: Technical and Philological Aspects!!
- Pipeline of characters and scripts
- Rejected scripts and characters
- Published Errata ✓

✖

2. Status Quo

- What is already available?
- What has been accomplished?
- A few appetizers

Russian Translit.

Transliteration for historical letters:

Θɵ: Ṫṫ (Irish Gaelic)
Vv: Ẏẏ ✓ (no info given)

(Latin Extd-Additional)

Russian Translit.

′

″

Miscellaneous phonetic modifiers

02B9 ′ MODIFIER LETTER PRIME
- primary stress, emphasis
- transliteration of mjagkij znak (Cyrillic soft sign: palatalization)

→ 0027 ' apostrophe
→ 00B4 ´ acute accent
→ 02CA ˊ modifier letter acute accent
→ 0301 ◌́ combining acute accent
→ 0374 ʹ greek numeral sign
→ 2032 ′ prime

02BA ″ MODIFIER LETTER DOUBLE PRIME
- exaggerated stress, contrastive stress
- transliteration of tverdyj znak (Cyrillic hard sign: no palatalization)

→ 0022 " quotation mark
→ 030B ◌̋ combining double acute accent
→ 2033 ″ double prime

✓

(Spacing Modifiers)

Macedonian Translit.

1E30 Ḱ LATIN CAPITAL LETTER K WITH ACUTE
≡ 004B K 0301 ◌́
1E31 ḱ LATIN SMALL LETTER K WITH ACUTE
• Macedonian transliteration
≡ 006B k 0301 ◌́

✓

01F4 Ǵ LATIN CAPITAL LETTER G WITH ACUTE
≡ 0047 G 0301 ◌́
01F5 ǵ LATIN SMALL LETTER G WITH ACUTE
• Macedonian and Serbian transliteration
≡ 0067 g 0301 ◌́

✗

✓

(Latin Extd-A, B)

➔

Cassubian

(Didn't have a single CodePage)

Kaszëbsczé abecadło:

aA ąĄ ãÃ bB cC dD eE éÉ ëË

fF gG hH iI jJ kK lL łŁ mM

nN ńŃ oO òÒ óÓ ôÔ pP rR

sS tT uU ùÙ wW yY zZ żŻ

✓

Nasal Vowels

Ąą Ęę Įį Ųų– Polish orthogr., Lithuanian

Ǫǫ – Sami, Old Icelandic

ǪǬ – Old Icelandic ✓

ǫ̆ ę̄ – not precomposed

3. Missing Pieces

Oversights, mistakes, problems, missing characters, encoding problems etc.

Štokavian Accents

	rising	falling
long	´	^
short	`	~

Vowels: a e i o u r

Štokavian Accents

á é í ó ú ŕ • Latin 1

à è ì ò ù _ • Latin 1 ✗

â ê î ô û r̂ • Extended B

ǎ ě ǐ ǒ ǔ ř • Extended B

ȓ missing from UC!

Štokavian Accents

random examples...

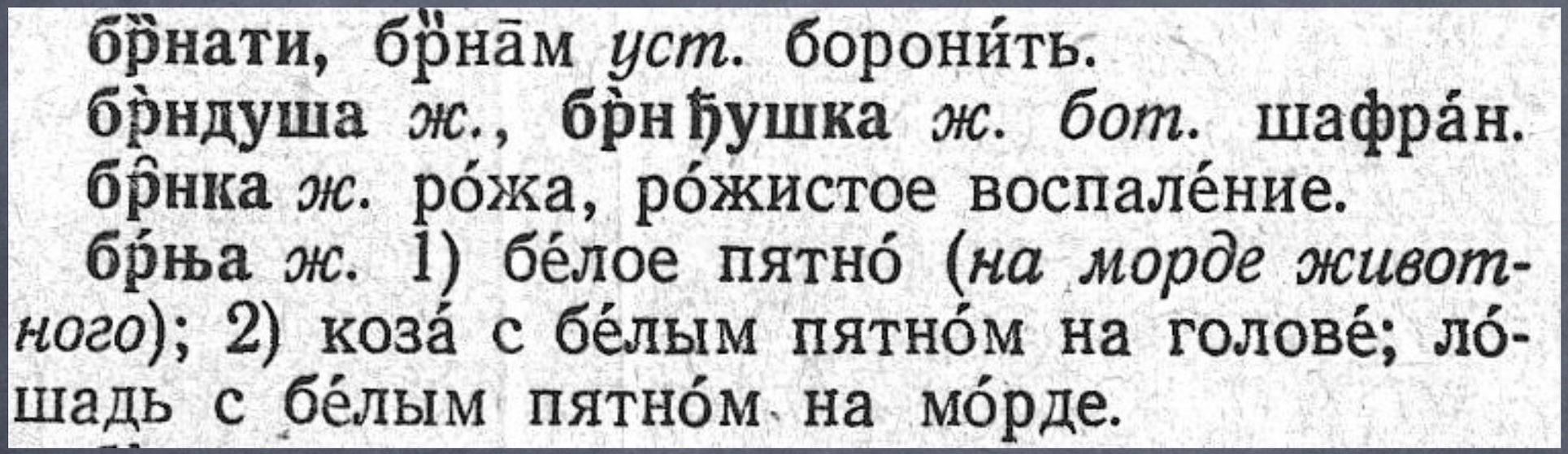
Sr̀bija

hr̀vatski

	rising	falling
long	´	^
short	`	~

Štokavian Accents

Accents in Serbian...



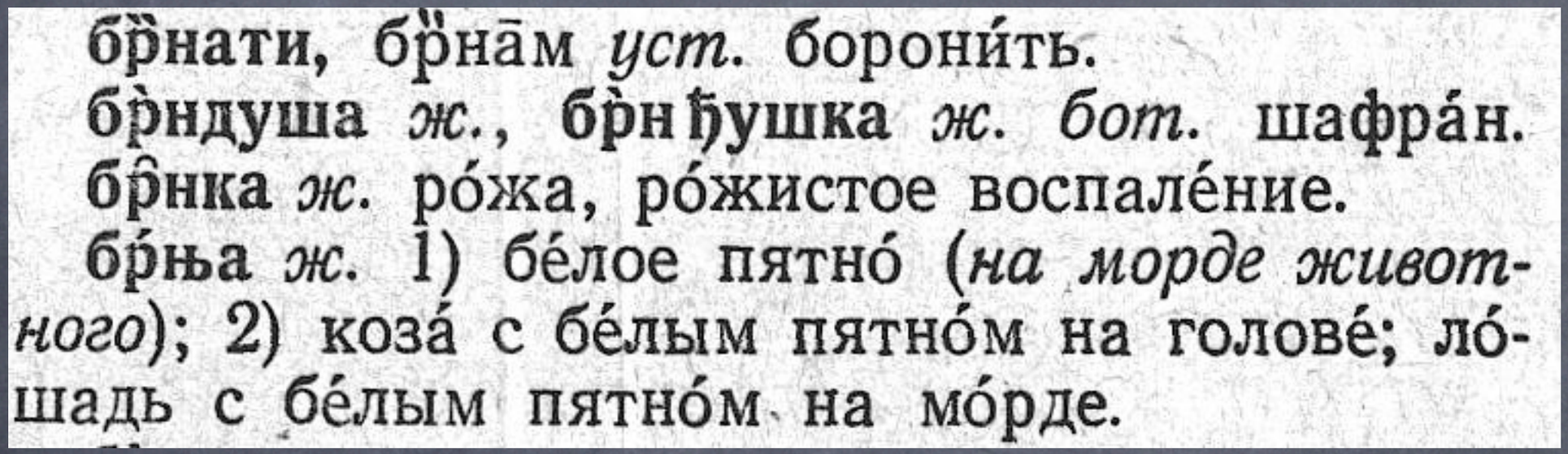
бр̏нати, бр̏нāм *уст.* бороня́ть.
бр̀ндуша *ж.*, бр̀нћушка *ж. бот.* шафра́н.
бр̂нка *ж.* ро́жа, ро́жистое воспале́ние.
бр́ња *ж.* 1) бе́лое пятно́ (*на морде животного*); 2) коза́ с бе́лым пятно́м на голове́; ло́шадь с бе́лым пятно́м на мо́рде.

...same as in Croatian!

И.И. Толстой (сост.), Сербскохорватско-русский словарь.
Изд. 3-е, исправл. и доп. Москва 1970.

Štokavian Accents

Accents in Serbian...



бр̏нати, бр̏нāм *уст.* бороня́ть.
бр̀ндуша *ж.*, бр̀нћушка *ж. бот.* шафра́н.
бр̂нка *ж.* ро́жа, ро́жистое воспале́ние.
бр́ња *ж.* 1) бе́лое пятно́ (*на морде животного*); 2) коза́ с бе́лым пятно́м на голове́; ло́шадь с бе́лым пятно́м на мо́рде.

....same as in Croatian!

...but not available in UC

Štokavian Accents

	rising	falling
long	´	̑
short	̀	̏

Latin: 23 of 24

á é í ó ú ŕ • Latin 1
à è ì ò ù _ • Latin 1
â ê î ô û r̂ • Extended B
ầ ề ì ồ ừ r̂̀ • Extended B

Cyrillic: nothing so far

✖

Croatian Digraphs

matching Serbian Cyrillic

uhorn	Upsilonafr	Vhook	Yhook	yhook	Zstroke	zstroke	Ezh	Ezhrevers	ezhrevers	ezhtail	twostroke	Tonefive	tonefive	glottalinve	wynn	
ư	Ʊ	Ʋ	Ƴ	ƴ	Ƶ	ƶ	Ʒ	Ƹ	ƹ	ƺ	ƻ	Ƽ	ƽ	ƾ	ƿ	
clickdental	clicklatera	clickalveo	clickretrot	DZcaron	Dzcaron	dzcaron	LJ	Lj	lj	NJ	Nj	nj	Acaron	acaron	Icaron	
ǀ	ǁ	ǂ	ǃ	DŽ	Dž	dž	LJ	Lj	lj	NJ	Nj	nj	Ǎ	ǎ	Ǐ	✓
icaron	Ocaron	ocaron	Ucaron	ucaron	Udieresisn	udieresisn	Udieresisa	udieresisa	Udieresisc	udieresisc	Udieresisgr	udieresisgr	eturned	Adieresisn	adieresisn	
ǐ	Ǒ	ǒ	Ǔ	ǔ	Ǖ	ǖ	Ǘ	ǘ	Ǚ	ǚ	Ǜ	ǜ	ǝ	Ǟ	ǟ	
Adotmacro	adotmacro	AEmacron	aemacron	Gstroke	gstroke	Gcaron	gcaron	Kcaron	kcaron	Oogonek	oogonek	Oogonekma	oogonekma	Ezhcaron	ezhcaron	
Ǡ	ǡ	Ǣ	ǣ	Ǥ	ǥ	Ǧ	ǧ	Ǩ	ǩ	Ǫ	ǫ	Ǭ	ǭ	Ǯ	ǯ	
jcaron	DZ	Dz	dz	Gacute	gacute	Hwair	Wynn	uni01F8	uni01F9	Aringacute	aringacute	AEacute	aeacute	Ostrokeacu	ostrokeacu	
ǰ	DZ	Dz	dz	Ǵ	ǵ	Ƕ	Ƿ	Ǹ	ǹ	Ǻ	ǻ	Ǽ	ǽ	Ǿ	ǿ	✓
Adblgrave	adblgrave	Ainvertedb	ainvertedb	Edblgrave	edblgrave	Einvertedb	einvertedb	Idblgrave	idblgrave	Iinvertedb	iinvertedb	Odblgrave	odblgrave	Oinvertedb	oinvertedb	
Ȁ	ȁ	Ȃ	ȃ	Ȅ	ȅ	Ȇ	ȇ	Ȉ	ȉ	Ȋ	ȋ	Ȍ	ȍ	Ȏ	ȏ	

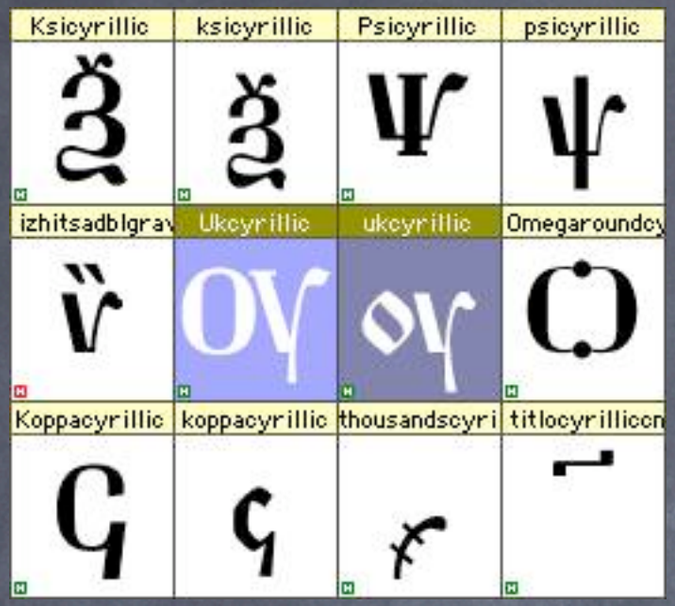
(Latin-Extended B)

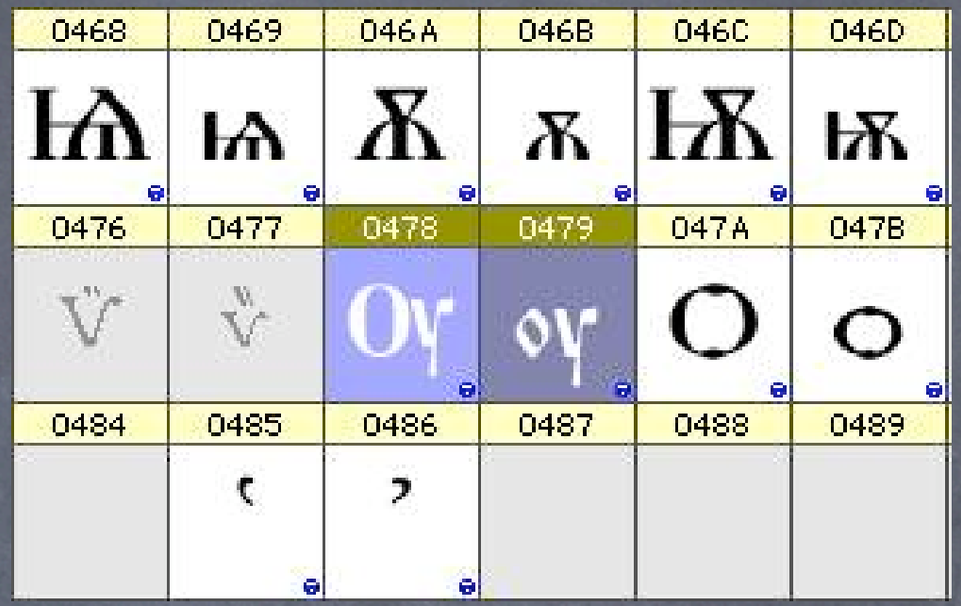
Cyrillic ‘UK’

two variants:
ligature or digraph

(Cyrillic Hist.)

Cyrillic OU/ou





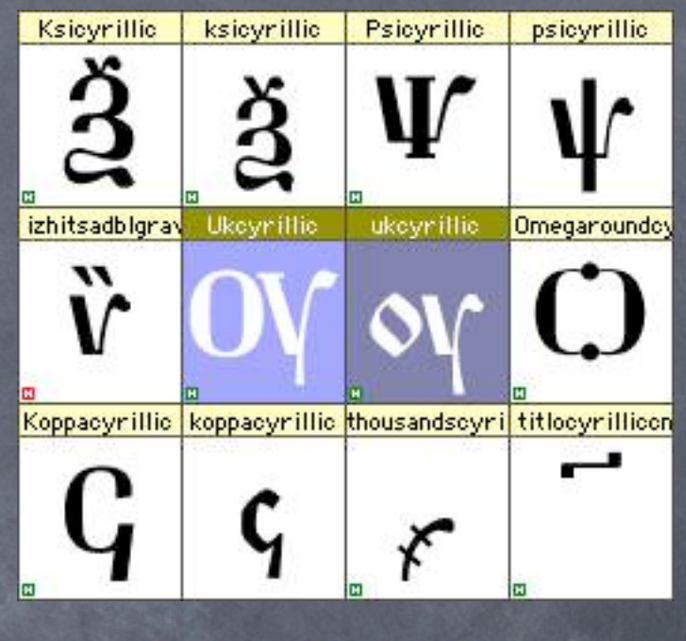
three forms needed

– just as for the Croatian digraphs!

UPPER CASE	MIXed	lower case
OY	Oy	oy

Cyrillic OU/ou

3 forms are needed;
but…
if only 2 are available,
they must be OU/ou,
not Ou/ou, because…



Cyrillic OU/ou

3 forms are needed;
but…
if only 2 are available,
they must be OU/ou,
not Ou/ou, because…

this doesn't look too good

✗ ОуГОДИТИ

оуГОДИТИ

Cyrillic OU/ou

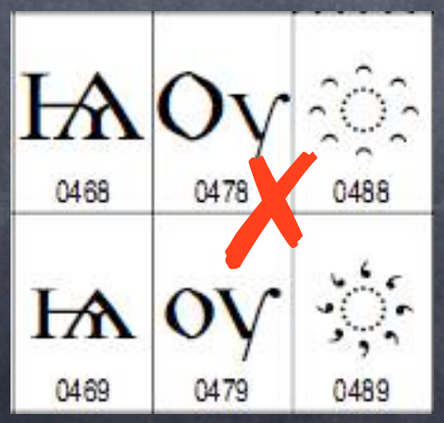
Mistake in Unicode 4.1 docs:

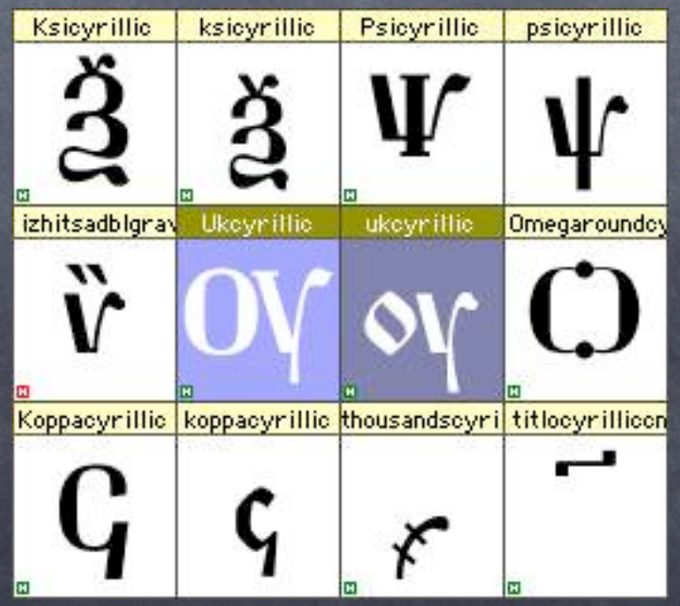
0478 Оу CYRILLIC CAPITAL LETTER UK

- basic Old Cyrillic uk is unified with CYRILLIC LETTER U

→ 0423 У cyrillic capital letter u

0479 оу CYRILLIC SMALL LETTER UK





Russian Phonetics

just added in UC 4.1:

ᵫ	ꝭ
1D6B	1D7B
ᵬ	ᵵ
1D6C	1D7C
ᵭ	ᵽ
1D6D	1D7D

ꝭ LATIN SMALL CAPITAL LETTER I WITH STROKE

- used with different meanings by Americanists and Oxford dictionaries

ᵵ LATIN SMALL LETTER IOTA WITH STROKE

- used by Russianists

Russian Phonetics

just added in UC 4.1:

ᵫ	ꝉ
1D6B	1D7B
ᵬ	ᵵ
1D6C	1D7C
ᵭ	ᵽ
1D6D	1D7D

ꝉ LATIN SMALL CAPITAL LETTER I WITH STROKE

- used with different meanings by Americanists and Oxford dictionaries

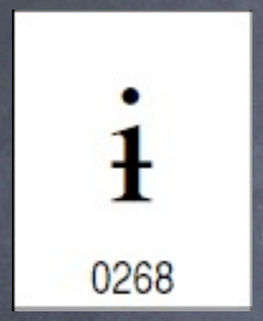
ᵵ LATIN SMALL LETTER IOTA WITH STROKE

- used by Russianists

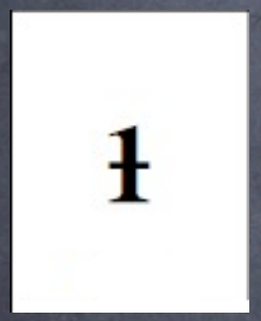
1) wrong
2) wrong

Russian Phonetics

really needed for [ы]:



✓ **needed and available**



✗ **needed but not available**

✗ **available but not needed**

Russian Phonetics

1		2		3
nach API	nach Avanesov	nach API	nach Avanesov	
[ı]	[и^{e}]	/e/	/e/	в лесу́
		/a/	/a/	пяти́, часте́й
		/o/	/o/	село́
[ɪ]	[ь]	/e/	/e/	лесово́д, о ле́се
		/a/	/a/	пятиле́тка, у́часть
		/o/	/o/	мо́рем, перенесена́
[ɨ]	[ы]	/i/	/и/	би́ты
[ɨ]	[ы^{e}]	/e/	/e/	шести́
[ʌ]	[ʌ]	/a/	/a/	дано́
		/o/	/o/	гото́в
[ə]	[ъ]	/e/	/e/	шестикла́ссник
		/a/	/a/	вы́дан, кни́га, пу́ля
		/o/	/o/	на́ год, ле́том, мо́рем

dotless ı, not iota!

K. Gabka (ed.), Einführung in das Studium der russischen Sprache. Phonetik und Phonologie. Düsseldorf 1975, 34 & 164.

Russian Phonetics

СТИХОТВОРЕНИЕ М. Ю. ЛЕРМОНТОВА „ПАРУС“

па́рус

б’и^{e}л’е̯́ьт [1] па́рус / лд’ино́къ̯ [2] /
ф-тума́·н’ь мо́·р’ъ / гълубо́м //
што́ и́ш’:ьт-он / ф-стрʌн’е́ дʌл’·о́къ̯ /
што́ к’и́нул-он / ф-крʌj·у́ рʌдно́м //

игра́·і̯·ут во́лны/ в’е́т’ьр с’в’и́ш’:ьт /
и-ма́·ч’тъ / гн’·о́^{т}цъ и-скрып’и́т [3] /
увы́ / о̀н-ш’:ӓ́с’т’и̯ъ [4] н’ь-и́ш’:ьт /
и-н’ь-лт-ш’:ӓ́с’т’и̯ъ [4] / б’и^{e}жы́т /

пʌд-н’и́м струj·а́ / с’в’и^{e}тл’е̯́і̯ лʌзу̯́·р’и /
нʌд-н’и́м / лу̯́·ч’-со́нцъ зълʌто́·і̯ /
л-о́н м’и^{e}т’е́жнъ̯і [5] / про́·с’ит бу́·р’и /
клг-бу́^{т}тъ [6] в-бу́·р’ъх је̯́с’т’ плко́·і̯ ///

Avanesov 1972

Cyrillic + Latin + Greek + Diacritics

Russian Phonetics

Cyrillic	абвгде…	✓
Latin/IPA	h i̯ j ə ʌ ↄ	✗
Greek	γ α	✓
Diacritics	ˆ ˊ ˋ • ¨ …	✗

Р.И. Аванесов, Русское литературное произношение, Москва 1972.

Russian Phonetics

Additional character 'Nasal': ◌̃

Самостоятельным, фонематическим признаком является „носовость вообще“ без локализации его образования. В словофонематической транскрипции носовую согласную фонему слабости по признаку дентальности-лабиальности обозначим знаком ◌̃. Вместе с тем отметим, что эта носовая фонема, слабая по признаку дентальности-лабиальности, одновременно является слабой и по признаку твердости-мягкости (так как перед [н'], [м'] она звучит мягко, а перед [н], [м]—твердо). Поэтому знак ◌̃ сопровождается цифрой 1 (справа ниже буквы), указывающей на неразличение этой слабой фонемы также по признаку твердости-мягкости. Исходя из сказанного приведенные выше слова в словофонематической транскрипции могут быть записаны следующим образом: |в'и◌̃$_1$т$_2$|, |ба◌̃$_1$т$_2$|, |ба́◌̃$_1$дα|, |кαма́◌̃$_1$д$_2$α|, |в'и́◌̃$_1$т'ик$_2$|, |ба́◌̃т'ик$_2$|; |ра́◌̃$_1$пα|, |ла́◌̃$_1$пα|, |α◌̃$_1$ба́р|, |бо́◌̃$_1$бα|, |ба◌̃$_1$б'и́т'$_2$|.

✗

Р.И. Аванесов, Русская литературная и диалектная фонетика, Москва 1974, 50.

Bulgarian Phonetics

Reviving OCS characters:

За означаване на дълга пауза в края на изречението: и дъждове #.

ѵ За означаване на африкат *дж:* [ѵоп], [ѵуѵѐ], [ѵас].

s За означаване на африкат *дз:* [sъ̀нкам], [sифт].

l За означаване на изговора на средно [л] пред гласна [lѐсно].

Граматика на съвременния български книжовен език. Том I. Фонетика. София 1982, 29

s = dz ✓

ѵ = dž ✗

Polish Orthography

breue longum groſſum molle

a *aa* *b* *b́*

c has quinque differencias in Polonico idiomate habet

c *k* *g* *cz* *ch*

per ſe molle breue longum durum molle

d *dz* *e* *ee* *ff* *f*

per ſe inproprie breue longum durum molle

g *ǵ*[87] *i* *ij* *l* *ĺ*

groſſum molle groſſum molle breue longum

m̨ *m* *n̨* *n* *o* *oo*

durum molle per ſe per ſe

ṕ *p* *q* *R r*

S in lingua Polonorum iſtas in ſe ſex differencias habet

ſ *s |* *ſſz* *ſch* *ſz* *zz*

breue longum

t *u* *uu* cum perdit vim vocalitatis,

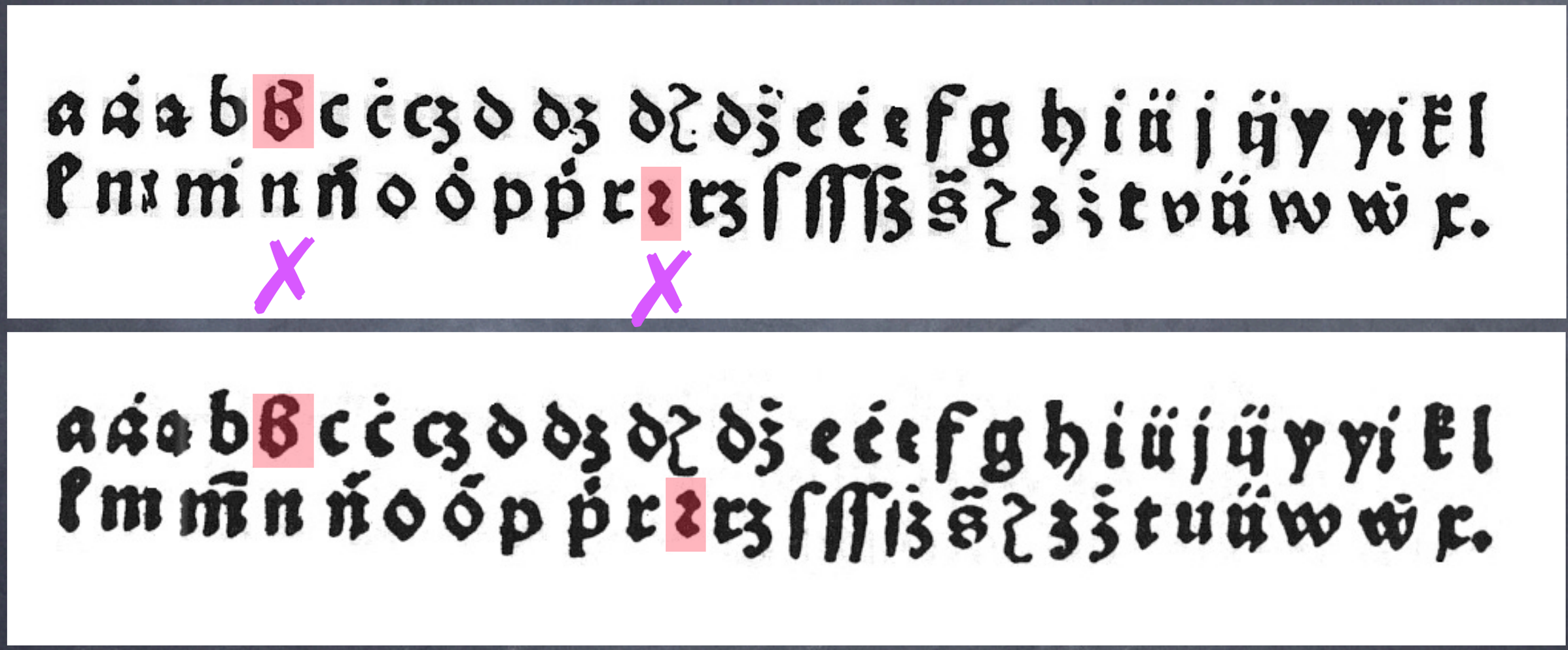
has tres habet differencias, que differencie patent in abece-

dario Polonorum, ſcilicet adaam bił etc.

v *v́* *w* *x* *y ||*

Jakub Parkosz, Traktat o ortografii polskiej.
Warszawa 1985, 78. (Original: ca. 1440)

Polish Orthography



St. Murzynowski, Orthographia Polska.
Königsberg 1551.

Polish Orthography

"r rotunda" - distinctive!

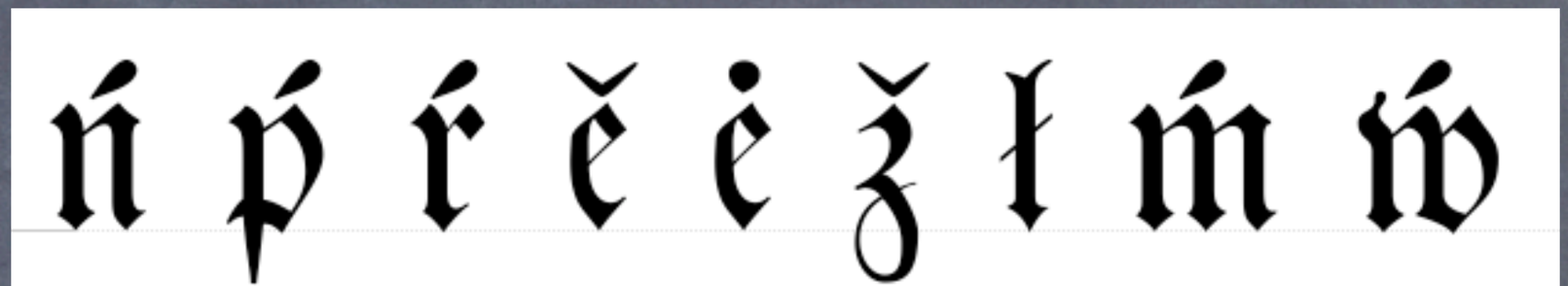
Um doppeldeutige Buchstabenverbindungen zu vermeiden, hat Murzynowski das Zeichen ꝛ eingeführt, da er die Verbindung *rz* = *r* von *r+z/r+ž* unterscheiden wollte in Beispielen wie *mierzy/mier-zi, skarzą/skarżą*. Der Buchstabe ꝛ wurde sehr häufig auch vom Schreiber des *Puławer Psalters* verwendet.

St. Murzynowski, Orthographia Polska. Königsberg 1551.

✗

Sorbian Orthography

Some Sorbian Characters:



✓

✗

(Latin Extd.-A, Extd. Additional)

Sorbian Orthography

Fabricius 1706	Fryco 1796	Zwahr 1847	Tešnaŕ-Šwjela	Časopis M. S.	Mein Alphabet	Ober-sorbisch	Alt-slovenisch
p	p	p	p	*p*	*p*	*p*	p
r	r ꝛ	r	r	*r*	*r*	*r*	r
—	sch — ſch	—	—	*ŕ*	—	*ŕ*	—
ß (ſſ)	ß	ss	ß	*s*	*s*	*s*	s
ſch	ßch	sch	ſch	*š*	*ś*	*ś*	š
ſch	ſch	schj	ſch	*ś*	*ś*	—	—
t	t	t	t	*t*	*t*	*t*	t

K.E. Mucke, Historische und vergleichende Laut- und Formenlehre der niedersorbischen (niederlausitzisch-wendischen) Sprache. Leipzig 1891 (Reprint 1965,18)

Sorbian Orthography

Typographic challenge: Design

S with stroke; r rotunda

Czech Orthography

Latin, Medieval German etc.

v = u = [u]

Czech Orthography

Until 1849:
initial [u] = v; [v] = w
Needed:

ý

ýdolj “valley”

✗

J. Schröpfer, Hussens Traktat „Orthographia Bohemica“ […], Wiesbaden 1968, 82, Fn. 25.

Czech Orthography

Needed for Czech: Ý ‘ýdolj’

Needed for German: Ÿ ‘fÿhret’

✗

✖

Historical Cyrillic

The Basic Principle:

"The historical form of the Cyrillic alphabet is treated as a font style variation of modern Cyrillic because the historical forms are relatively close to the modern appearance" (UC book, ch. 7)

Historical Cyrillic

The Basic Principle:

"The historical form of the Cyrillic alphabet is treated as a font style variation of modern Cyrillic because the historical forms are relatively close to the modern appearance" (UC book, ch. 7)

- creates as many problems as it solves!

Historical Cyrillic

A Glimmer of Hope:

"If, at some future date, the old letterforms are adequately documented and the need for them demonstrated, then they can be added to this [the Cyrillic-Extended] block" (Unicode book v. 1, vol. 1, 45)

OCS, Old Russian

Application more or less trivial…

Ѧ ≈ А

Ѣ ≈ Б

Щ ≈ Щ

etc.

OCS, Old Russian

functional identity

Г ≈ Љ

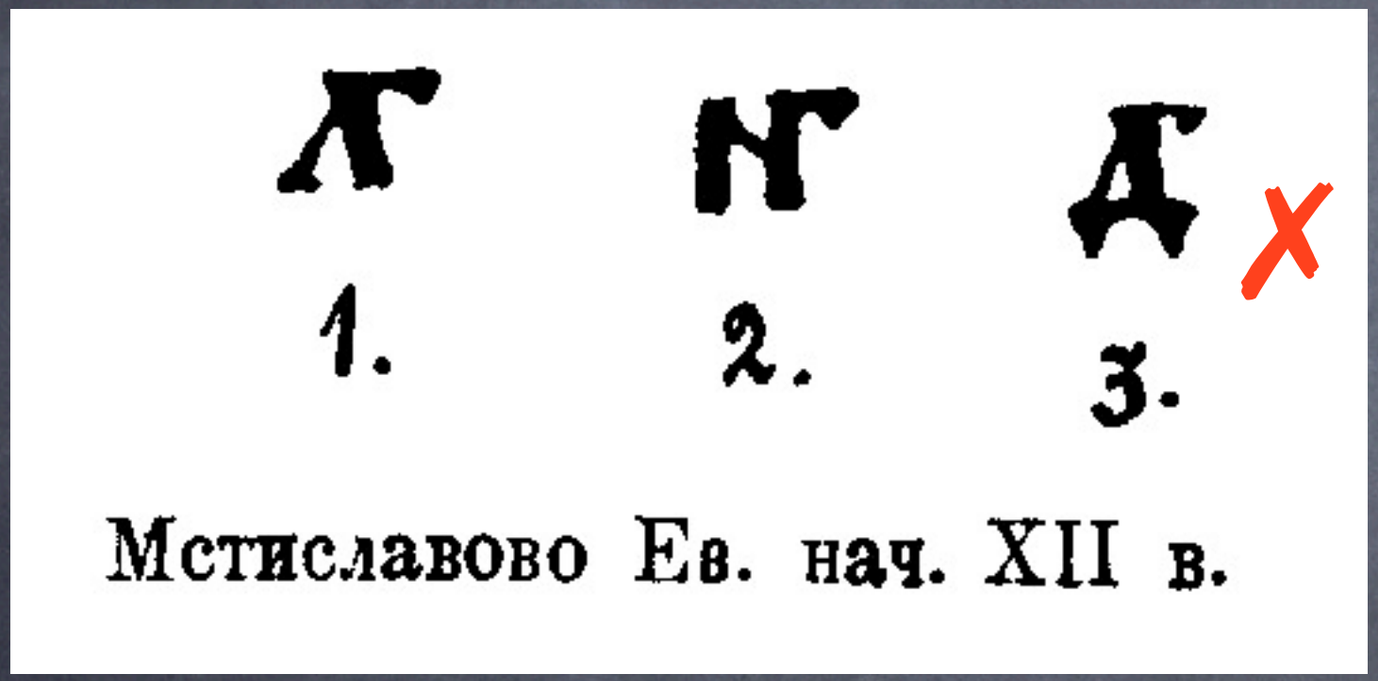
Нг ≈ Њ

Г̑ ≈ Ѓ

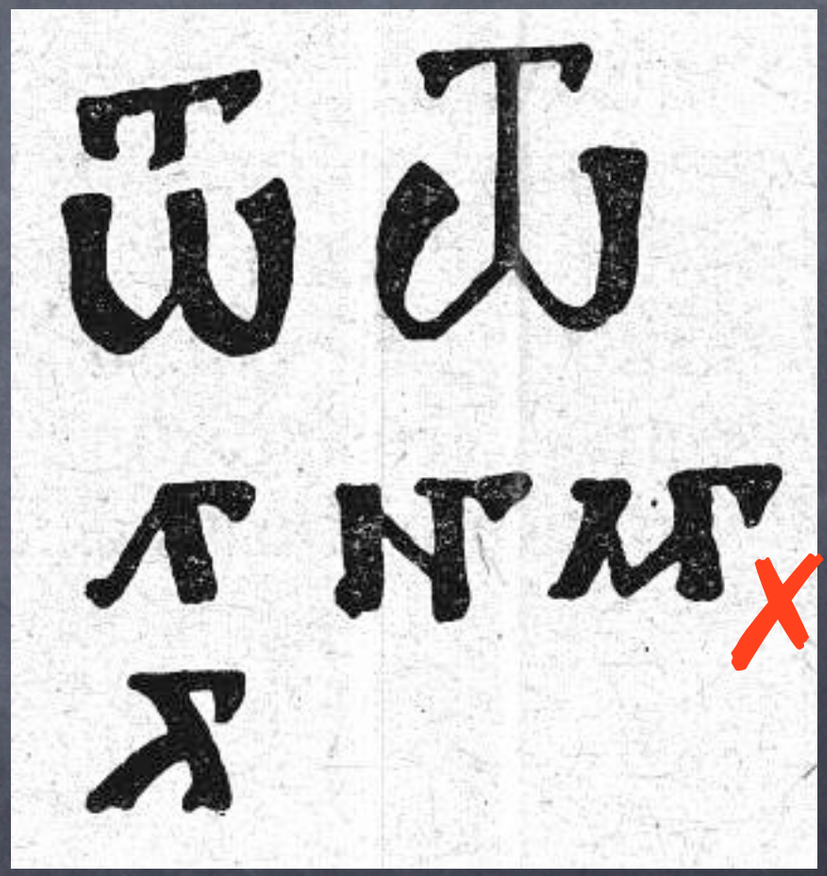
К̑ ≈ Ќ

OCS, Old Russian

but then we find...



А.И. Соболевский, Славяно-русская палеография. Изд. 2-е, СПб. 1908, 52.



В.Н. Щепкин, Русская палеография, Москва 1967, 111.

OCS, Old Russian

functional identity no correspondence

Л ≈ Љ Л ≈ ?

Н ≈ Њ

Д ≈ Ђ р́ ≈ ?

Ѓ ≈ Ѓ ѓ ≈ ?

Ќ ≈ Ќ х́ ≈ ?

OCS, Old Russian

best-known problem...

Ꙗ ≈ Я ?

Ѧ ≈ Ѧ ?

- Unicode does not have Ꙗ
- being added to 'private area'

Early Modern Russian

	[jo]	
18th century (AG-1802)	îo	✗
Karamzin 1797	ë	✓

variation: iô, jô, ию, iô…

Early Modern Russian

	[jo]	[ja]	
‘Old Form’	і̂о	ѧ	✗
‘Modern Glyph’	ё	я	

Unicode: status quo not satisfying; too important for private area!

Old Russian

Jotated Jat'

Separate glyphs or variants of jat?

Old Russian

Jotated Jat'

Separate glyphs or variants of jat?

Old Russian

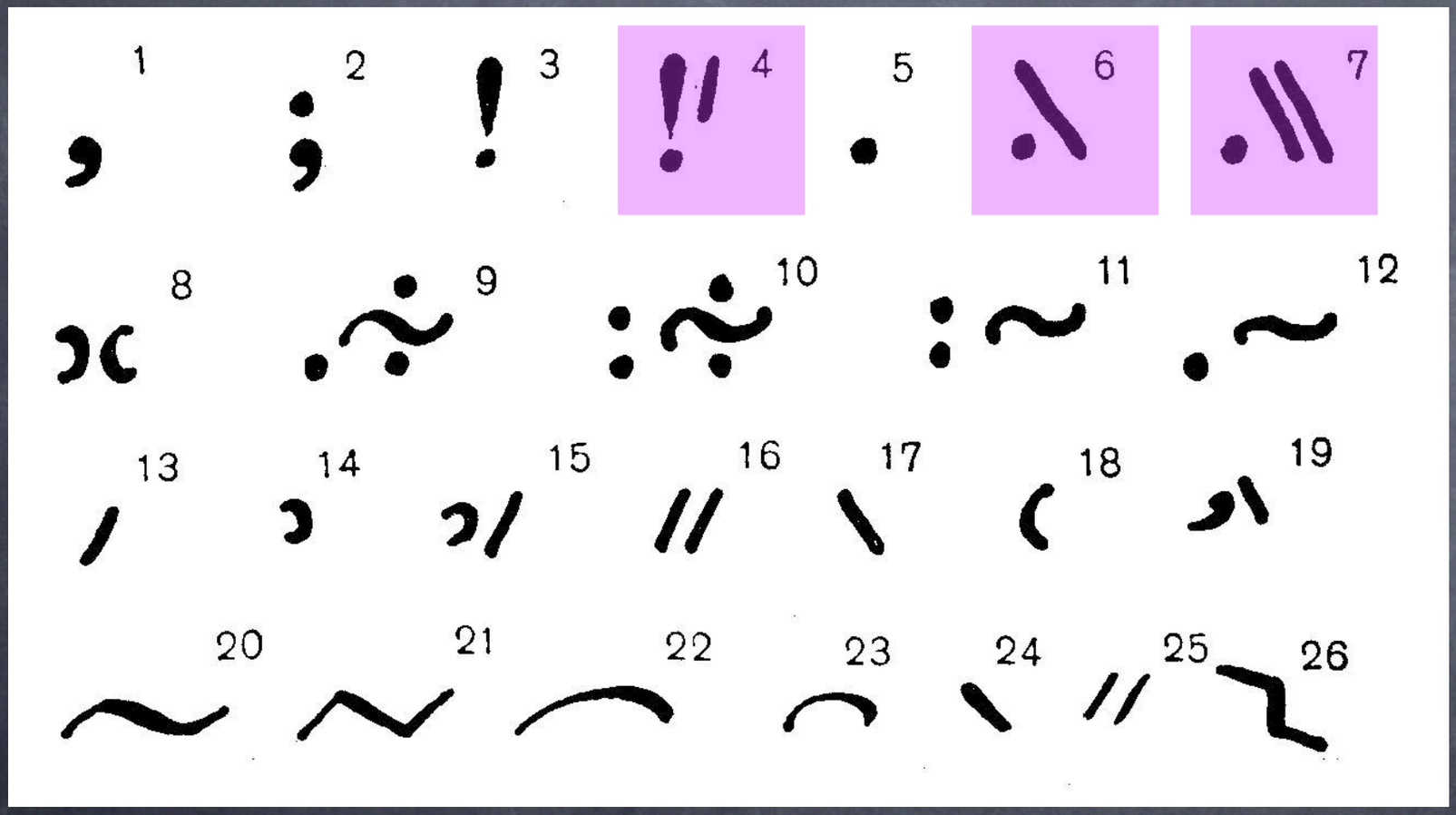
Not encoded yet:

Ѧ Ѩ Ѫ Ѭ

Ѯ Ѱ Ѳ Ѵ Ѡ Ѿ

Old Russian

znaki prepinanija 1

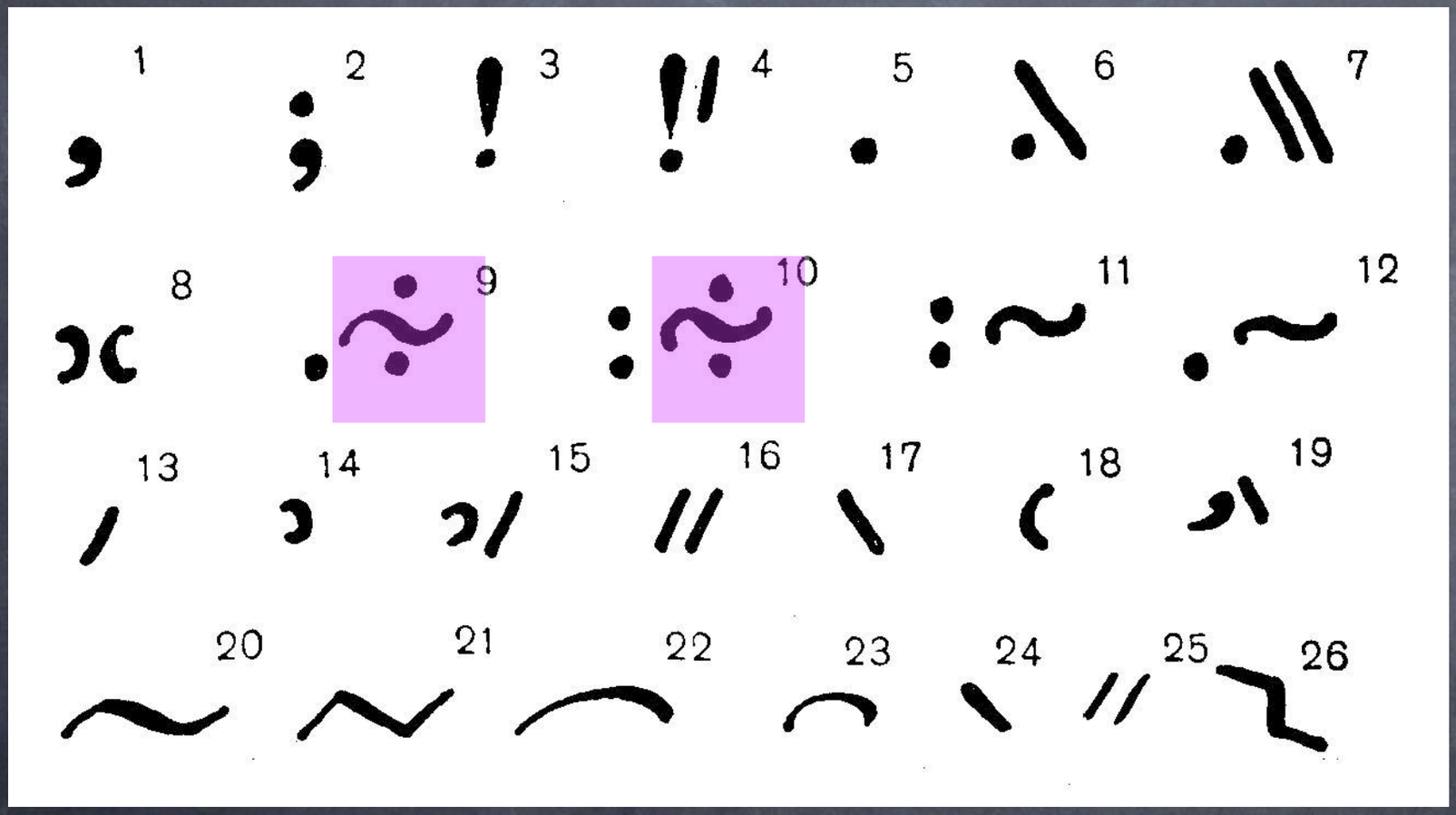


1 zapjataja
2 podstolija
3 pribyl'ca
4 pribavlenaja
5 točka
6 položitel'naja
7 kendema

Л.В. Черепнин, Русская палеография,
Москва 1956, 374-376.

Old Russian

znaki prepinanija 2



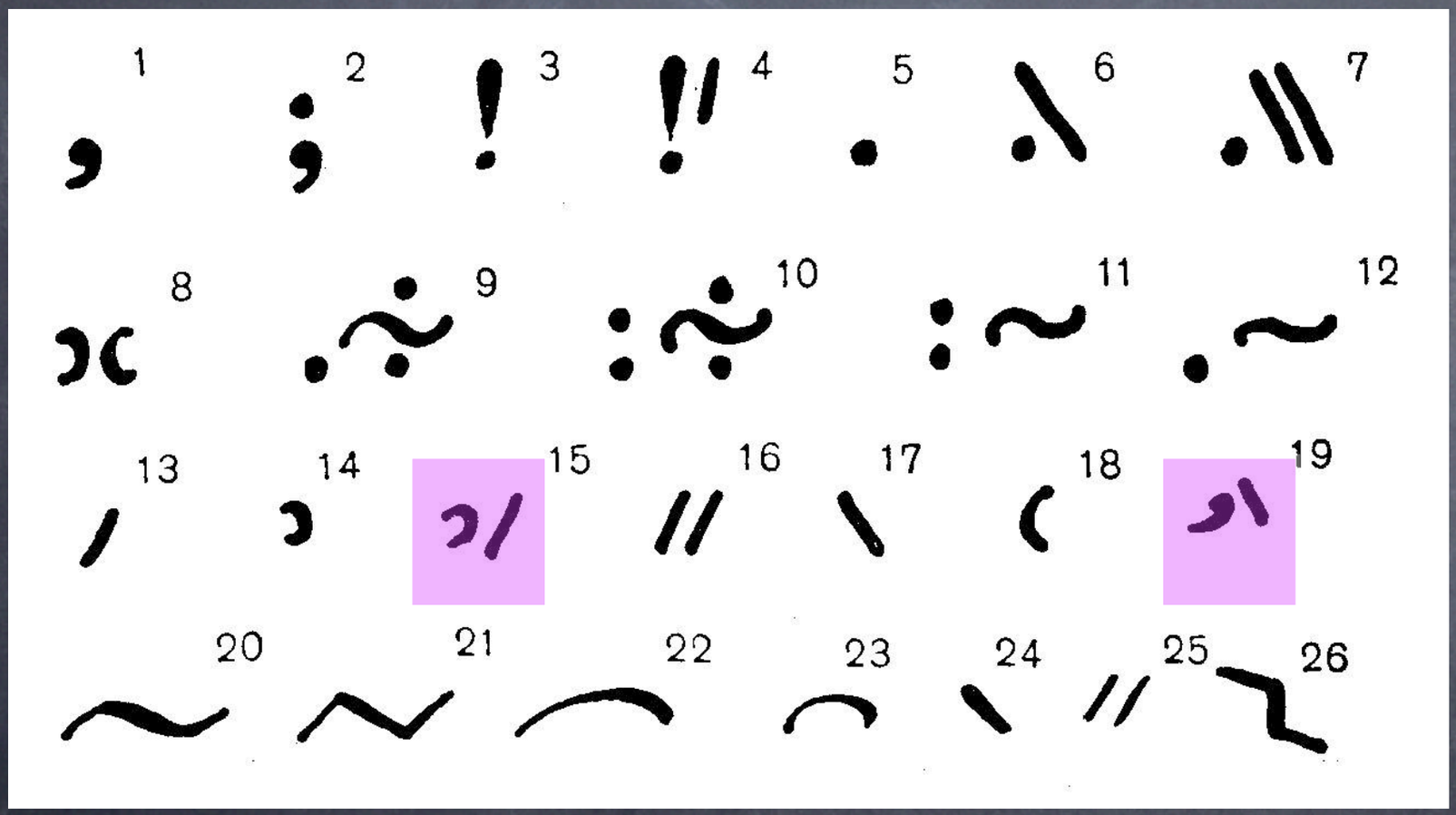
8 kavyka
9 slogija
10 statija
11 stišica
12 otrikal’

✗

Л.В. Черепнин, Русская палеография,
Москва 1956, 374-376.

Old Russian

udarenija i pridyxanija

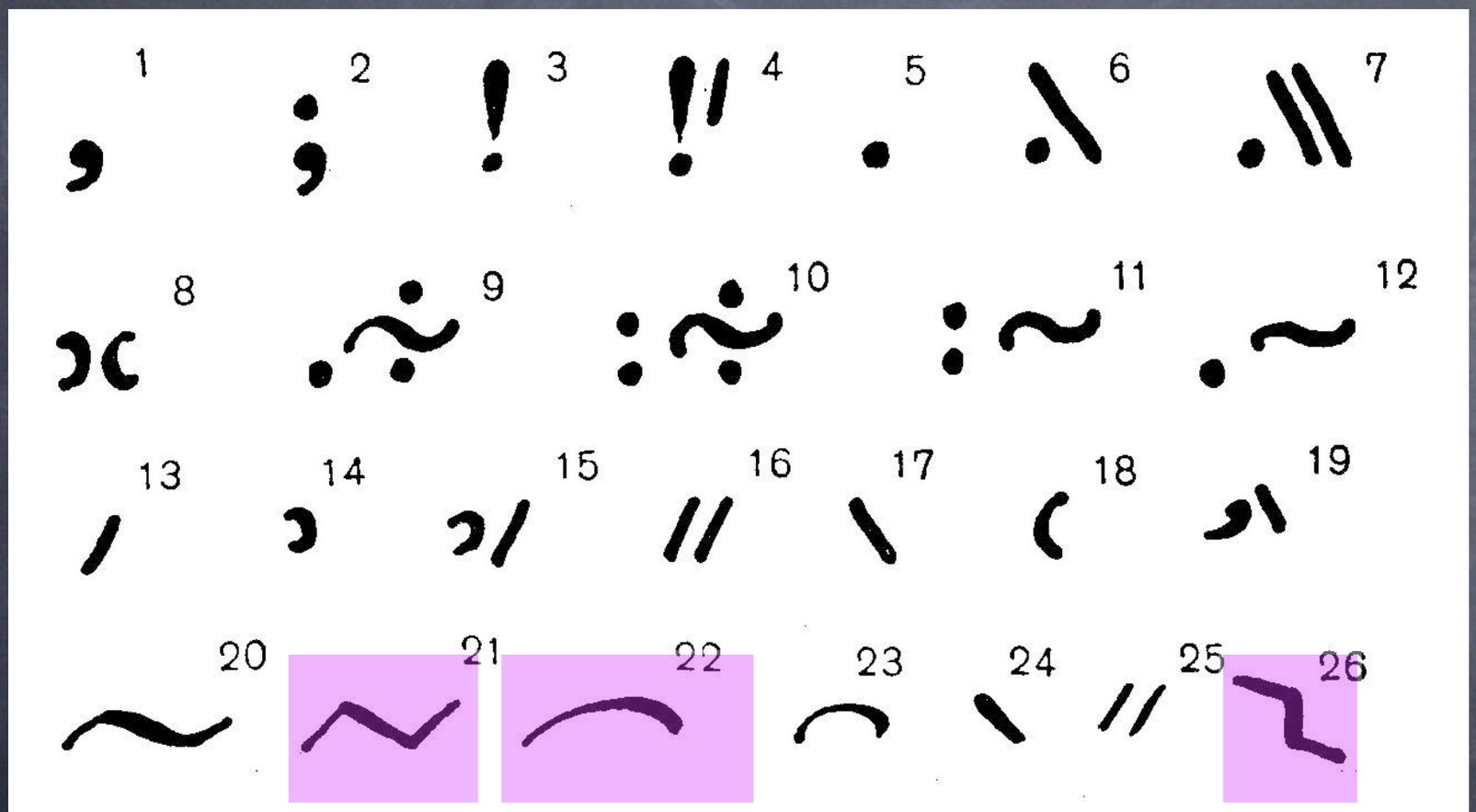


13 oksija
14 zvatelco/psili
15 iso/isso
16 okovavy
17 varija
18 dasija
19 apostrof ✗

Л.В. Черепнин, Русская палеография,
Москва 1956, 374-376.

Old Russian

Titlo & Co.



20 titlo
21 vzmet
22 pokrytie
23 kamora
24 stjaga
25 smyček
26 erok, ertica ✗

Л.В. Черепнин, Русская палеография,
Москва 1956, 374-376.

Old Russian

problem: paerok

1. знак ’ или ⁀ : мор’е, к’итъ, въпл⁀ь.

2. знак ’ или ‘ : м‘ного (мъного), д‘ва (дъва), в‘се (вьсе), ч’то (чьто).

3. титло ~, ⁀, ¯, ~ : сн̃ъ (сънъ), вл̃ка (владъка), іс̃ъ̃ (ісоусъ), гл̃а (глава), мц̑ь (мѣсяць).

Н.Г. Самсонов, Древнерусский язык.
Москва 1973, 57

Old Russian

problem: paerok

2. знак ' или ʻ : М'НОГО (МЪНОГО), Д'ВА (ДЪВА), В'СЕ (ВЬСЕ), Ч'ТО (ЧЬТО).

- has two basic forms
- stands between characters!
- is not a diacritic!

Old Russian

problem: paerok

2. знак ’ или ‘ : М‘НОГО (МЪНОГО), Д‘ВА (ДЪВА),
В‘СЕ (ВЬСЕ), Ч’ТО (ЧЬТО).

- has two basic forms
- stands between characters!
- is not a diacritic!

≠ UC+033E "comb. vertical tilde": ‘

Old Russian

part of the solution

02BC ʼ MODIFIER LETTER APOSTROPHE

= apostrophe

• glottal stop, glottalization, ejective

• spacing clone of Greek smooth breathing mark

• many languages use this as a letter of their alphabets

• 2019 ’ is the preferred character for a punctuation apostrophe

→ 0027 ' apostrophe

→ 0313 ◌̓ combining comma above

→ 0315 ◌̕ combining comma above right

→ 055A ՚ armenian apostrophe

→ 2019 ’ right single quotation mark

- letter!
- Mac./ Russ.
- one of two forms of paerok

‘apostrophe’ - Leskien

Old Russian

current Unicode solution?

	[jo] 0451	[ja] 044F	apostr. 02BC
‘Old Form’	іо̂	ꙗ	҇
‘Modern Glyph’	ё	я	’

Old Church Slavonic

2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	215A	215B	215C	215D	215E	215F
▒	▒	▒	1/3	2/3	1/5	2/5	3/5	4/5	1/6	5/6	1/8	3/8	5/8	7/8	1/
2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	216A	216B	216C	216D	216E	216F
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	L	C	D	M
2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	217A	217B	217C	217D	217E	217F
i	ii	iii	iv	v	vi	vii	viii	ix	x	xi	xii	l	c	d	m

Roman Numerals: Everything present ✗

Old Church Slavonic

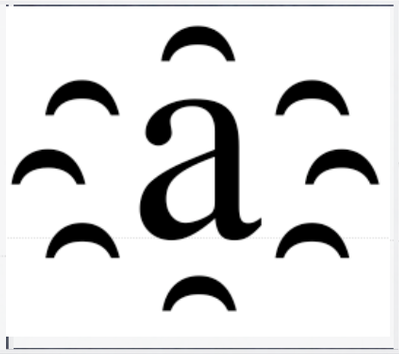
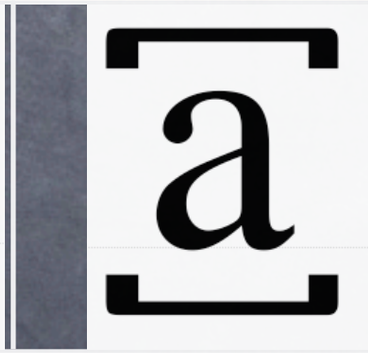
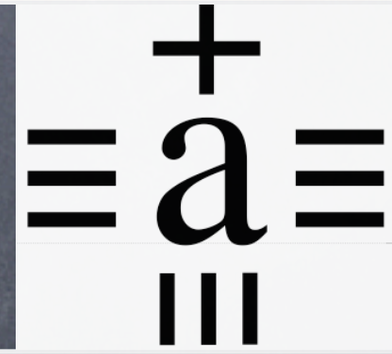
Roman Numerals: Everything present ✗

Slavic Numerals: parts only ✗

✖

Old Church Slavonic

Numbers, Numbers

10 T t'ma	100 T legion	1 Mio leodr	10 Mio voron	100 Mio koloda	1000 Mio t’ma tem
					

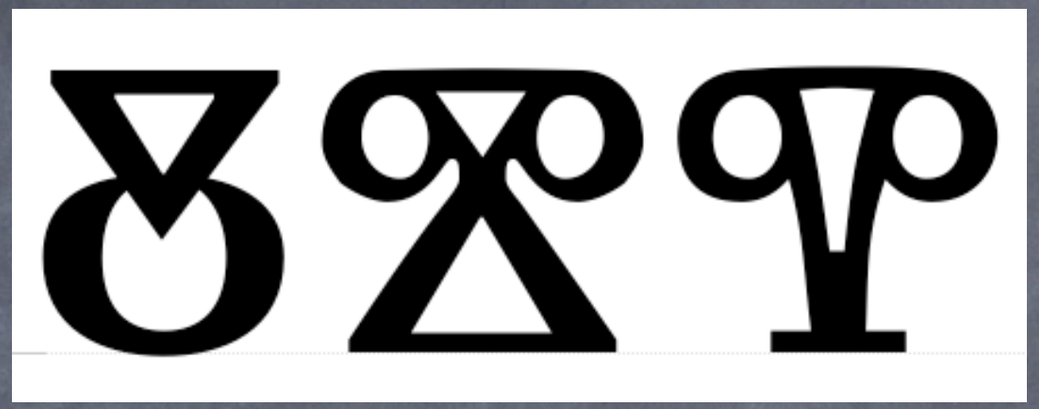
А.Х. Востоков, Грамматика церковно-словенскаго языка изложенная по дрейвнейшимъ онаго письменнымъ памятникамъ. СПб. 1863, 9 (Reprint Köln 1980)

Old Church Slavonic

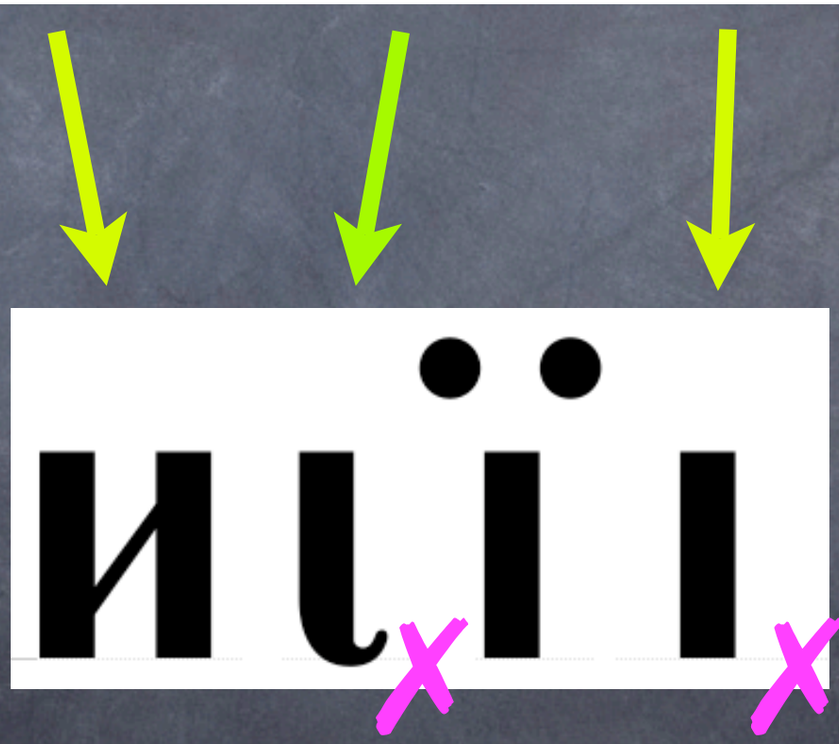
encoded so far...

10 T t'ma	100 T legion	1 Mio leodr	10 Mio voron	100 Mio koloda	1000 Mio t’ma tem
(x)	x	x			

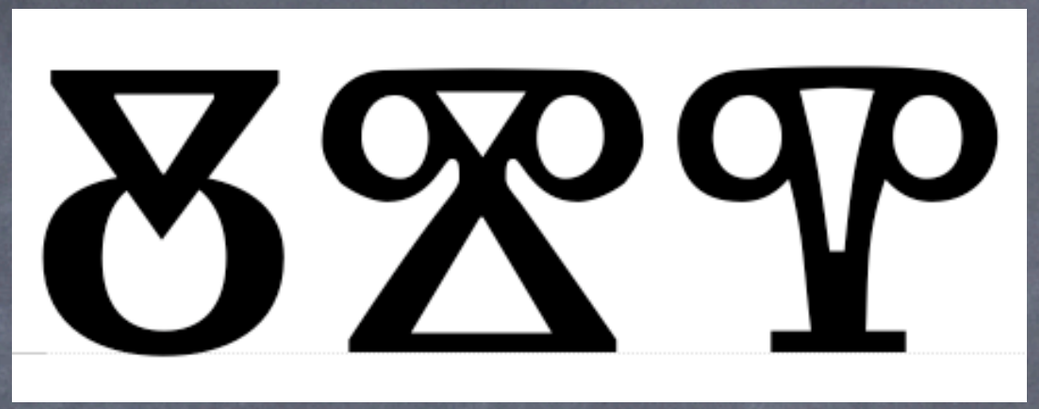
Glagolitic Translit.



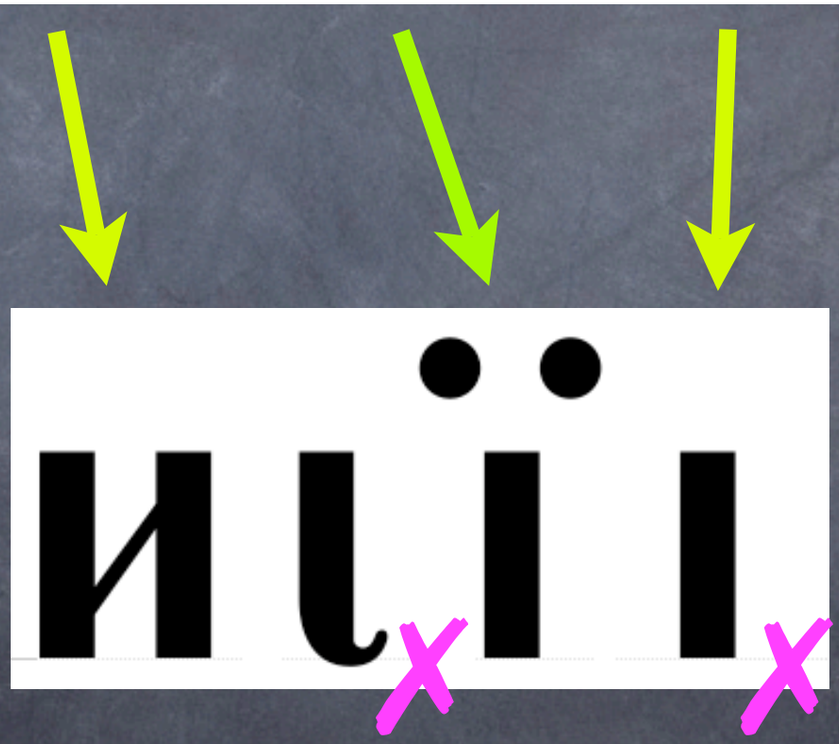
✓



Glagolitic Translit.

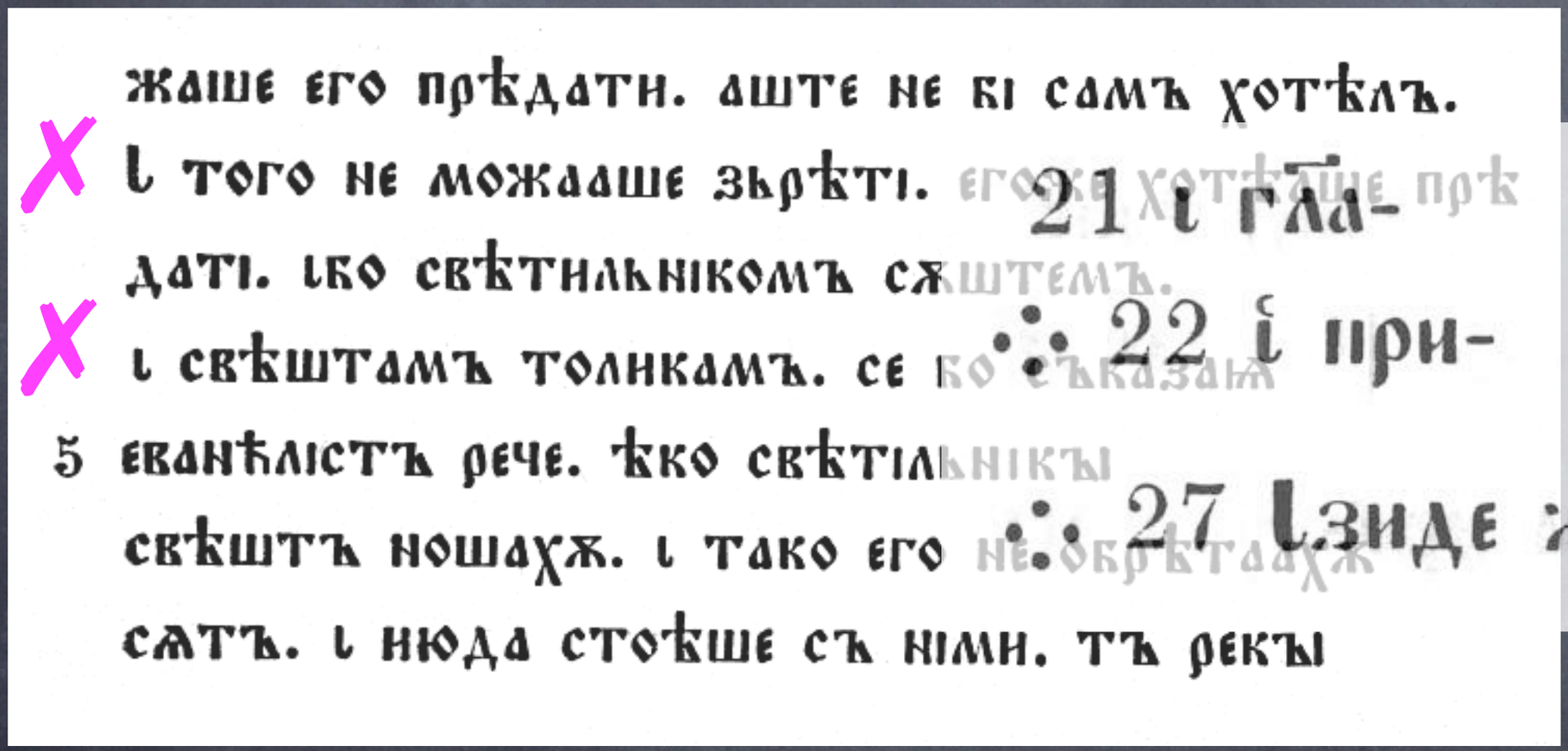


✓



Glagolitic Translit.

Samples: Cyrillic Iota



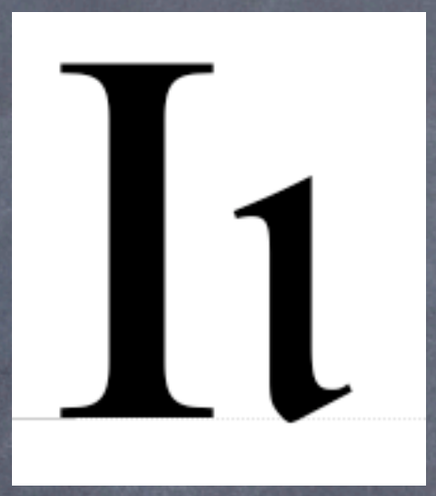
И.В. Ягичъ, Глаголическое письмо. В: Энциклопедія славянской филологіи, вып. 3, “Графика у славян”, СПб. 1911, 232.

Jagic, Edition of Codex Marianus, passim

Glagolitic Translit.

Design: Cyrillic Iota

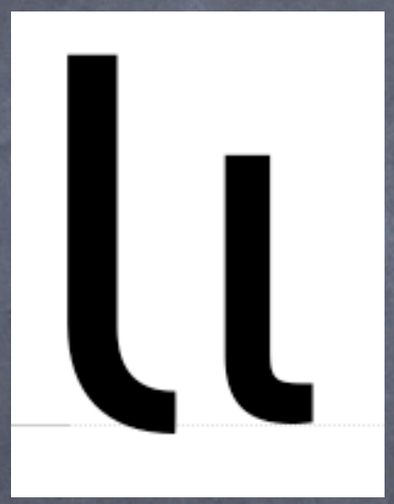
OCS
Black



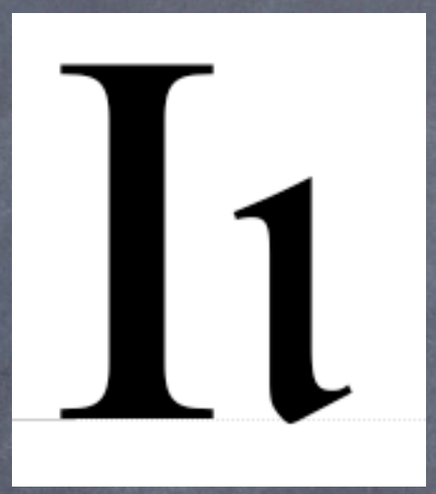
Greek
Serifed

Glagolitic Translit.

Design: Cyrillic Iota



Cyrillic
Sans
Roman



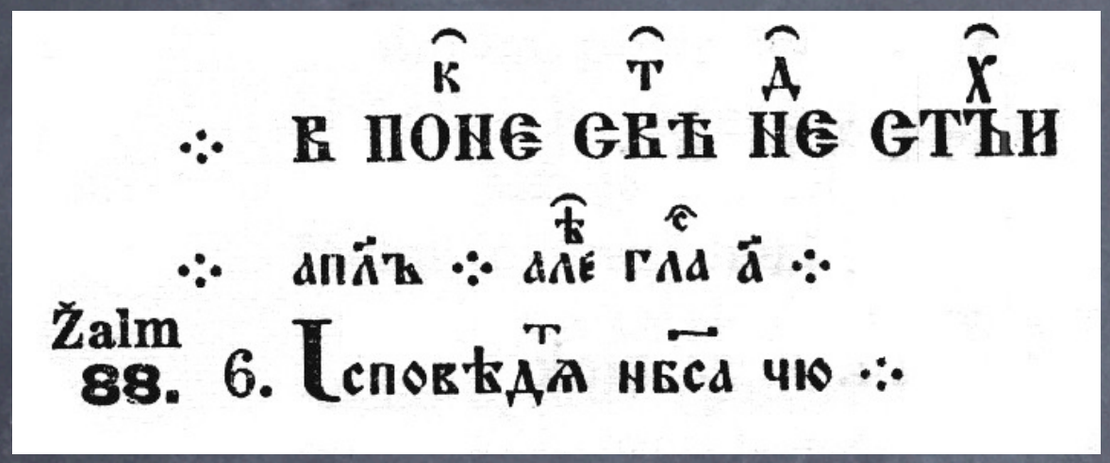
Greek
Serifed
Roman

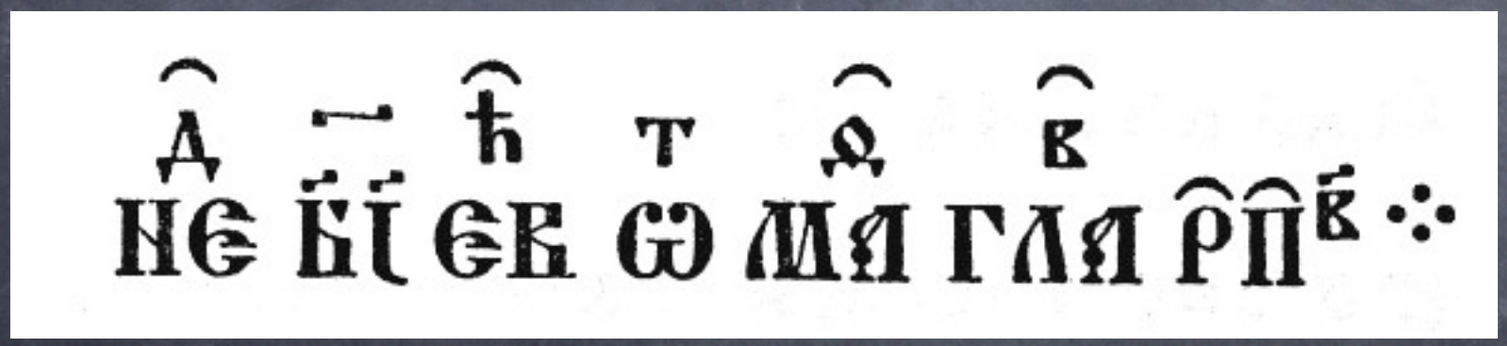
Cyrillic
Serifed
Roman

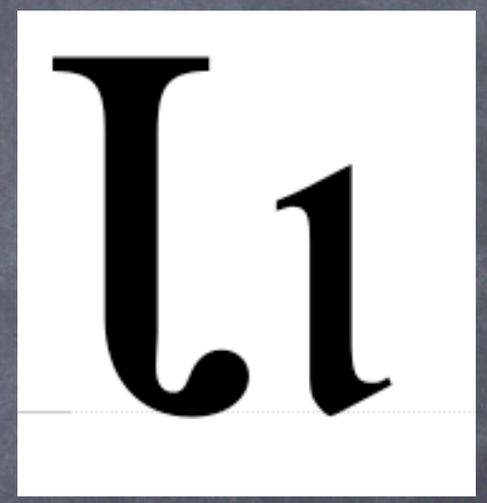
✗

Glagolitic Translit.

Design: Cyrillic Iota







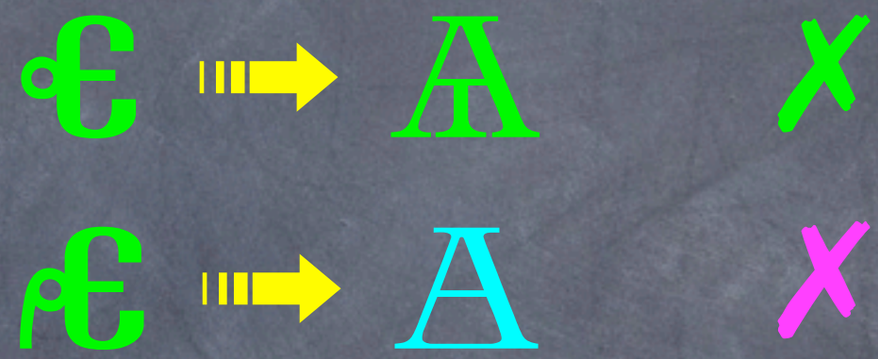
Evangeliarum Assemani. Codex Assemani 3. slavicus glagoliticus. Ediderunt Josef Vajs & Josef Kurz. Tomus II. Edidit Josef Kurz. Pragae 1955, 3, 89

Glagolitic Translit.

Transliteration of Nasals

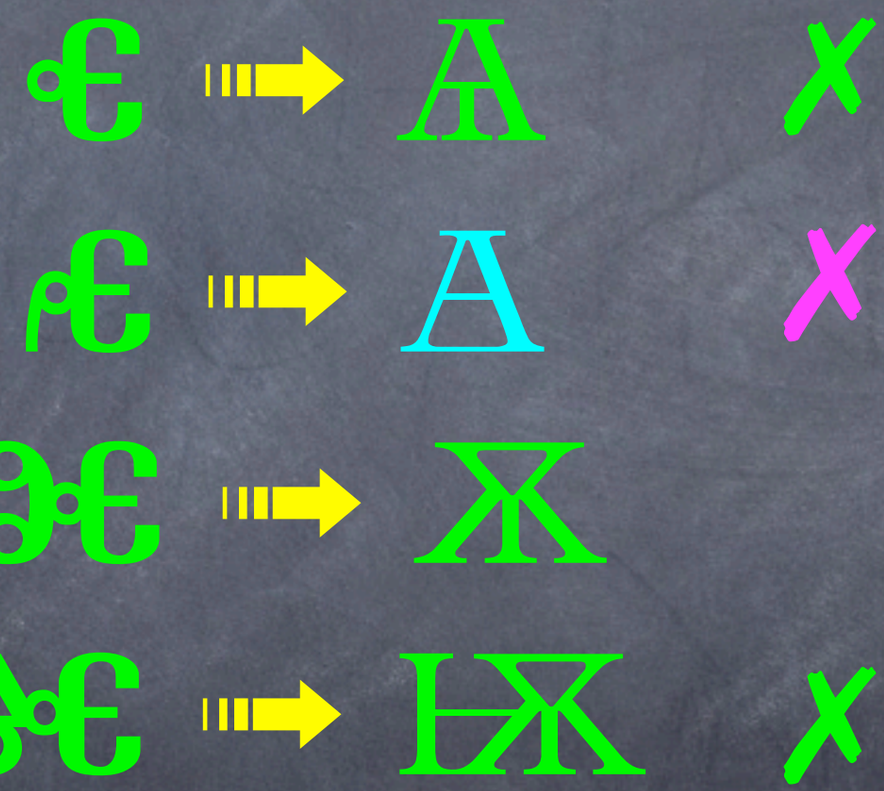
Glagolitic Translit.

Transliteration of Nasals



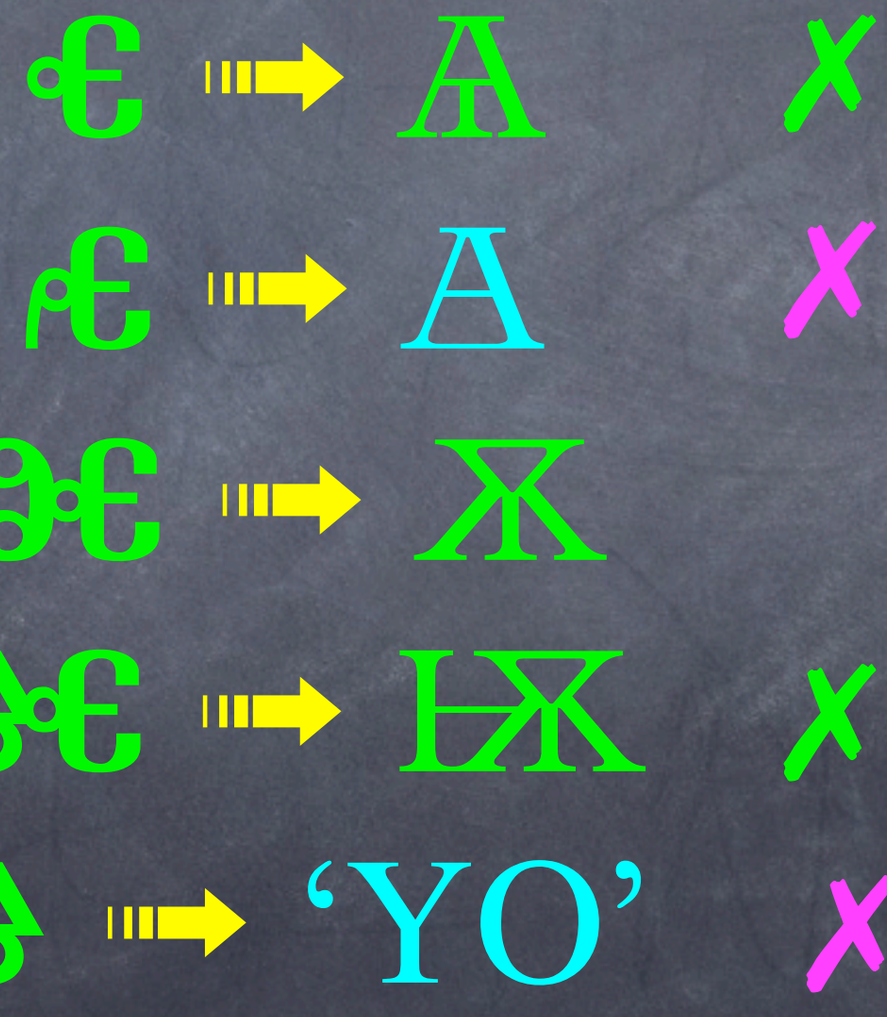
Glagolitic Translit.

Transliteration of Nasals



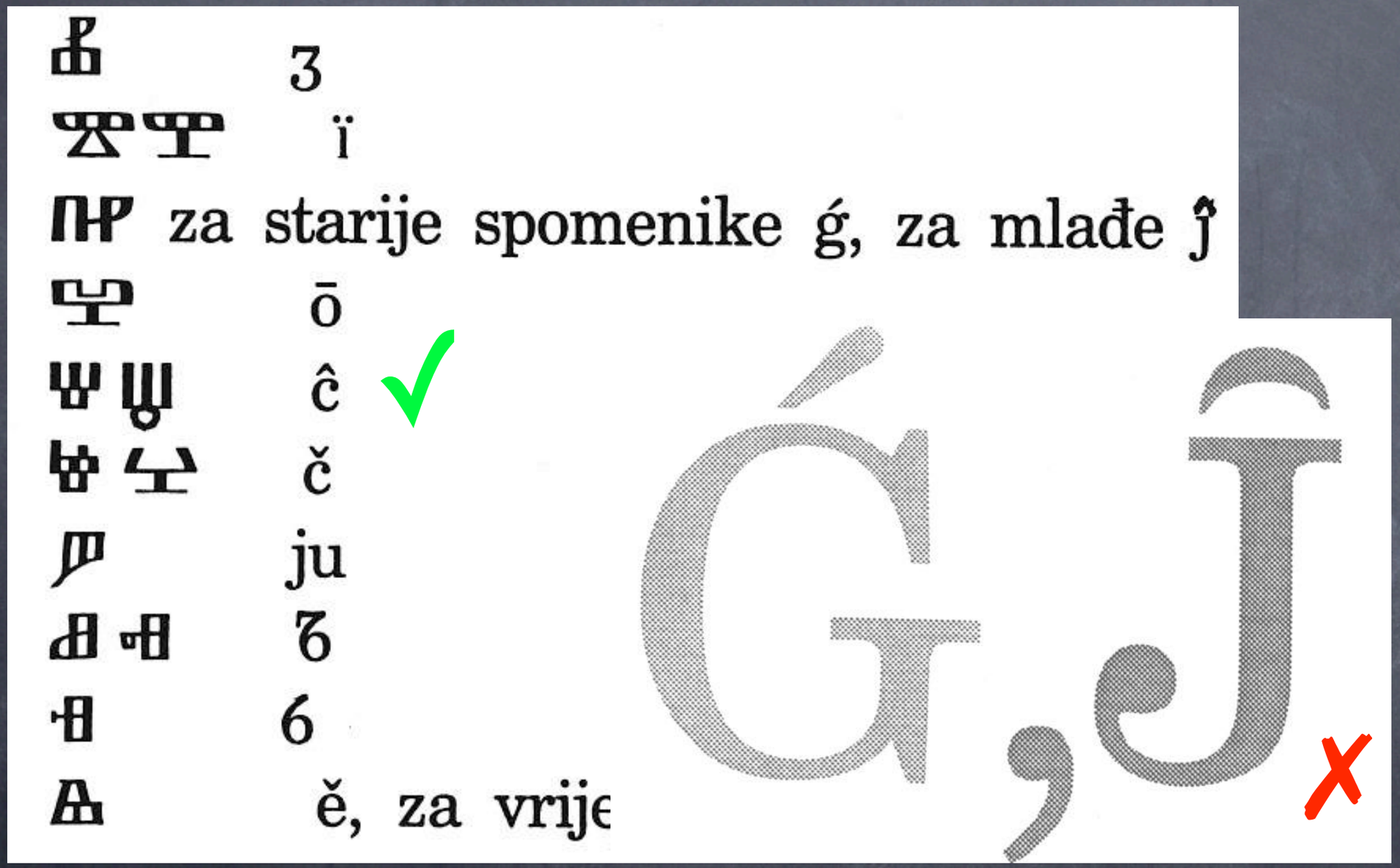
Glagolitic Translit.

Transliteration of Nasals



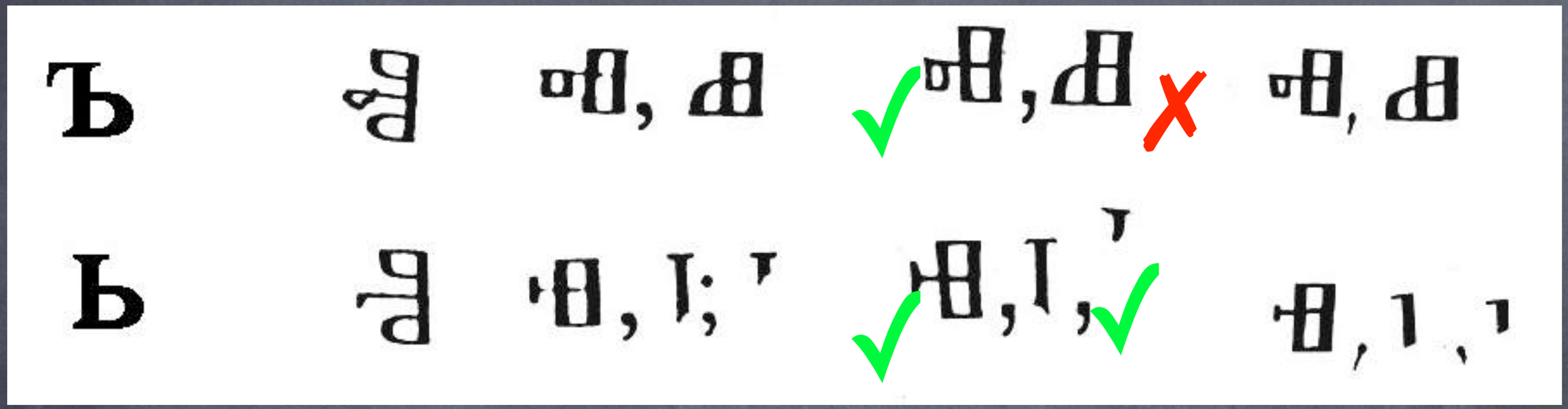
Glagolitic Translit.

Croatian Norm



J. Bratulić, Leksikon hrvatske glagoljice, Zagreb 1995, 98, 146.

Croatian Glagoljica



Different characters
or stylistic variants?

J. Bratulić, Leksikon hrvatske glagoljice, Zagreb 1995, 175.

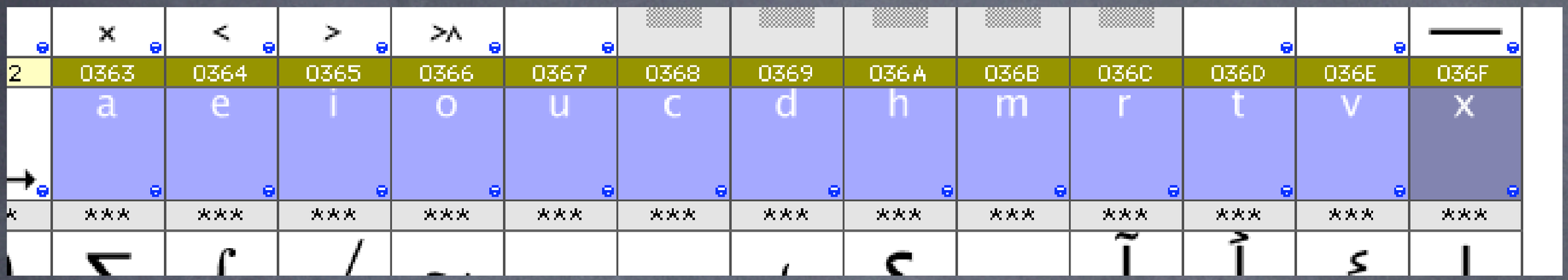
Bosančica

ⰰ	A
ⰱ	B
ⰲ	V
ⰳ	G
ⰴ, Ⱄⰴ, ⰵ	D
ⰵ	E
ⰶ, ⰶⰵ, ⰷ	Ž
ⰸ	DZ
ⰹ	Z
ⰻ	I
ⰼ, ⰽ	Đ, Ć (đerv)
ⰾ	K
ⰿ	L
ⰼⰿ	LJ
ⱀ	M
ⱁ	N
ⰽⱁ	NJ
ⱂ	O
ⱃ	P
ⱄ	R
ⱅ	S
ⱆ	T
ⱇ	U
ⱈ	F
ⱉ	H
ⱊ	OT
ⱋ	Ć, ŠT, ŠĆ
ⱌ	C
ⱍ	Č
ⱎ	Š
ⱏ	poluglas
ⱐ, ⱑ	JAT
ⱓ	JU

✗

✖

Superscripts



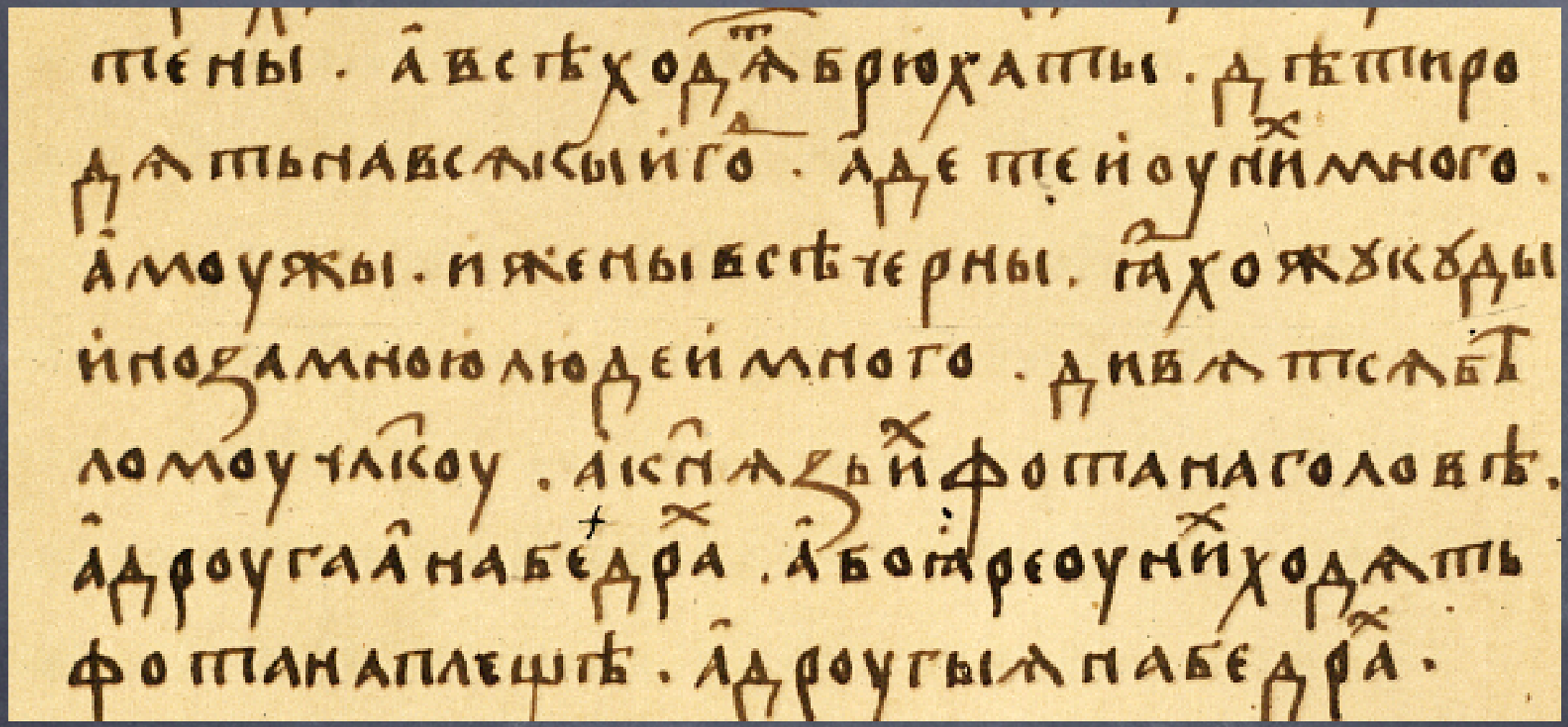
	x	<	>	>^									
2	0363	0364	0365	0366	0367	0368	0369	036A	036B	036C	036D	036E	036F
	a	e	i	o	u	c	d	h	m	r	t	v	x
*	***	***	***	***	***	***	***	***	***	***	***	***	***

Latin: Many Superscripts (cmb) for Medievalists (German!):
u͛ (=ü) a͛ (=ä) o͛ (=ö); v͛ etc.

✓

(Comb. Diacritical Marks)

Superscripts

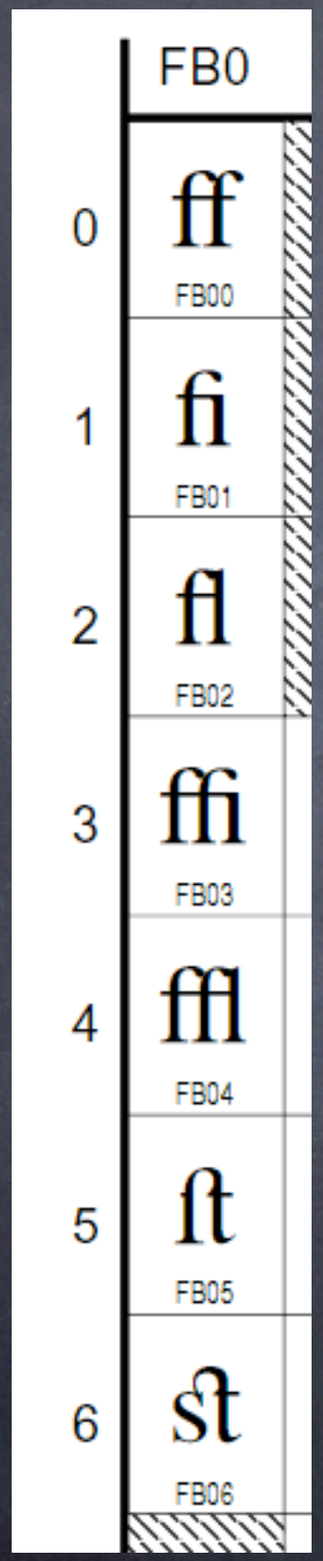


Cyrillic: No Superscripts at all in UC ✗

Афанасий Никитин, Хожение за три моря
Троицкий список

✖

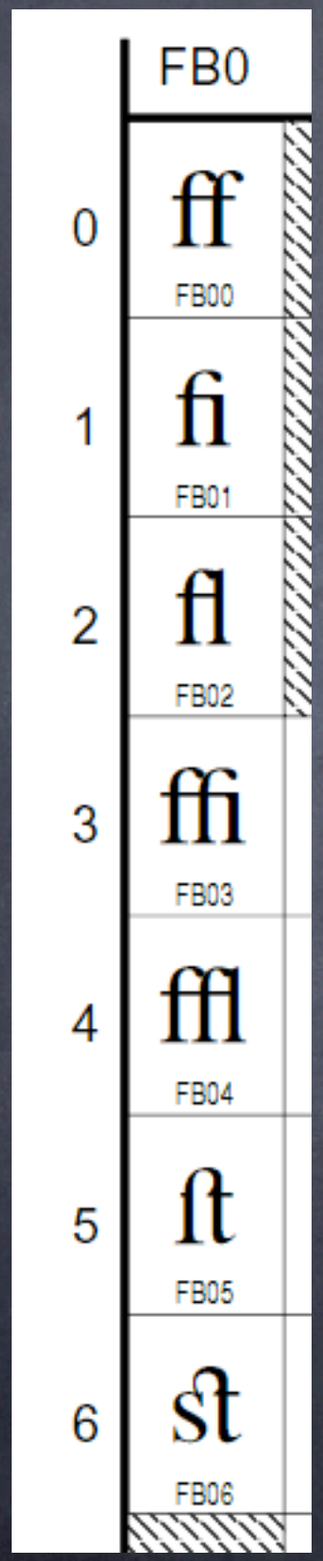
Ligatures



- "Presentation forms"
- some for Latin
- many for Arabic

✓

Ligatures



- "Presentation forms"
- some for Latin
- many for Arabic
- none for Cyrillic
- none for Glagolitic

Balkan Philology

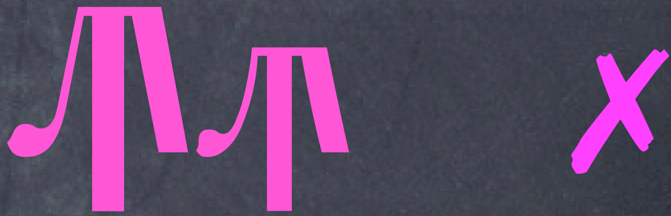
- OCS-Cyrillic for Romanian

LITERA Ѧ

Litera Ѧ, obținută prin modificarea grafică a literei Ѫ[78], în documentele muntenești ale secolului al XV-lea apare o singură dată, într-un act de întărire dat pentru mănăstirea Govora[79] de voievodul Radu cel Mare: **ѡт глав поіанє долє по ѦН**[80] **долинѹ Чорѫчєв** „din capul poienei, în jos, prin (pe în) valea Ciorîcei“ (1499 iul. 13, Arh. St. Buc., S. I., nr. 118).

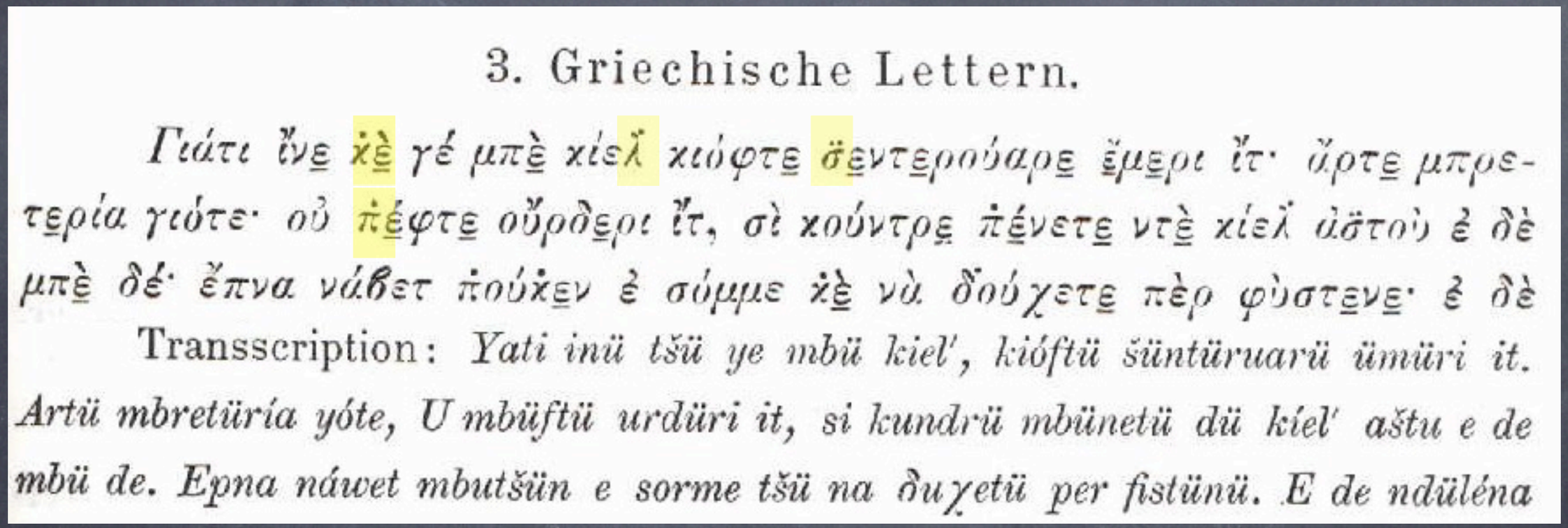
Deși atestată într-un cuvînt românesc, totuși unicul exemplu în care găsim această literă nu ne dă posibilitatea de a ne referi la valoarea sau valorile acestei litere în general[81]. În acest unic caz, grafia ne determină să afirmăm că Ѧ notează nazalitatea vocalei care precedă pe Н[82].

L. Djamo-Diaconiţă: Limba documentelor slavo-române emise în Ţara Românească în sec. XIV şi XV. Bucureşti 1971, 49



Balkan Philology

- Greek for Albanian



3. Griechische Lettern.

Γιάτι ἴνε ϰὲ γέ μπὲ ϰίελ ϰιόφτε σ̈ὲντεροὺαρε ἔμερι ἴτ· ἄρτε μπρε-τερία γιότε· οὺ πὲφτε οὔρδερι ἴτ, σὶ ϰούντρε πὲνετε ντὲ ϰίελ ἀσ̈τοὺ ἐ δὲ μπὲ δέ· ἔπνα νάβετ πούϰεν ἐ σύμμε ϰὲ νὰ δούχετε πὲρ φὺστενε· ἐ δὲ

Transscription: *Yati inü tšü ye mbü kiel', kióftü šüntüruarü ümüri it. Artü mbretüría yóte, U mbüftü urdüri it, si kundrü mbünetü dü kiel' aštu e de mbü de. Epna náwet mbutšün e sorme tšü na duχetü per fistünü. E de ndüléna*

Balkan Philology

- Greek for Albanian

3. Griechische Lettern.

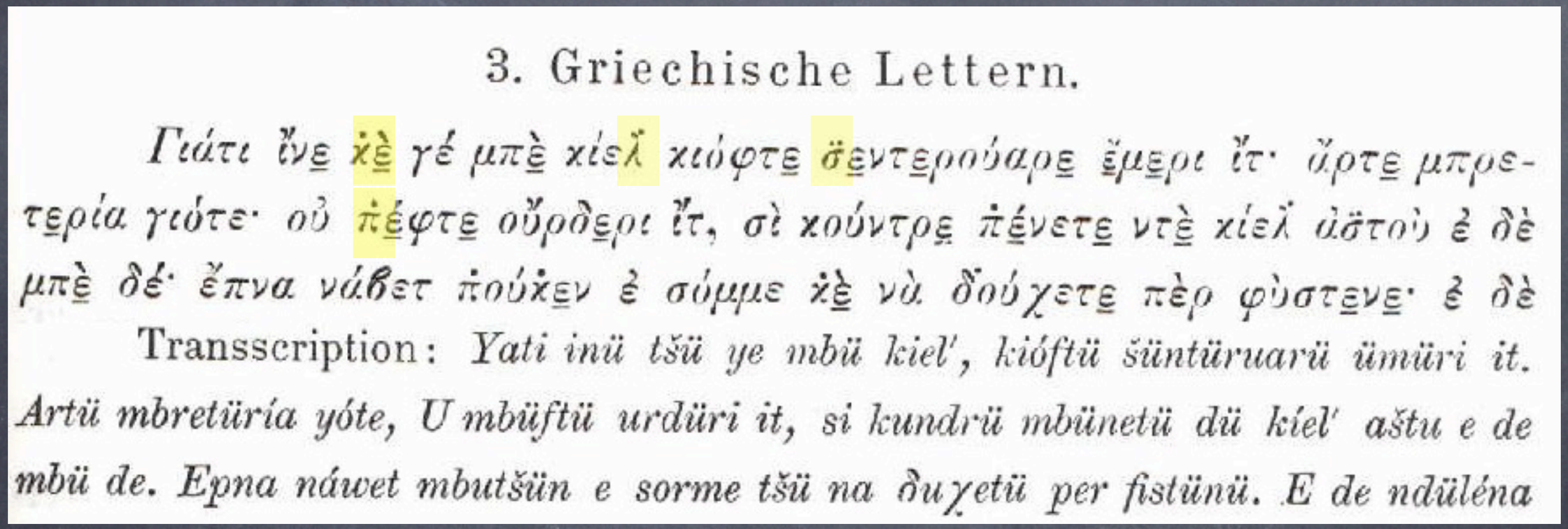
Γιάτι ἴνε ϰὲ γέ μπὲ ϰίελ ϰιόφτε σ̈ὲντεροὺαρε ἔμερι ἴτ· ἄρτε μπρε-τερία γιότε· οὺ πὲφτε οὔρδερι ἴτ, σὶ ϰούντρε πὲνετε ντὲ ϰίελ ἀσ̈τοὺ ἐ δὲ μπὲ δέ· ἔπνα νάβετ πούϰεν ἐ σύμμε ϰὲ νὰ δούχετε πὲρ φὺστενε· ἐ δὲ

Transscription: *Yati inü tšü ye mbü kiel', kióftü šüntüruarü ümüri it. Artü mbretüría yóte, U mbüftü urdüri it, si kundrü mbünetü dü kiel' aštu e de mbü de. Epna náwet mbutšün e sorme tšü na duχetü per fistünü. E de ndüléna*

- Greek for Macedonian…

Balkan Philology

- Greek for Albanian



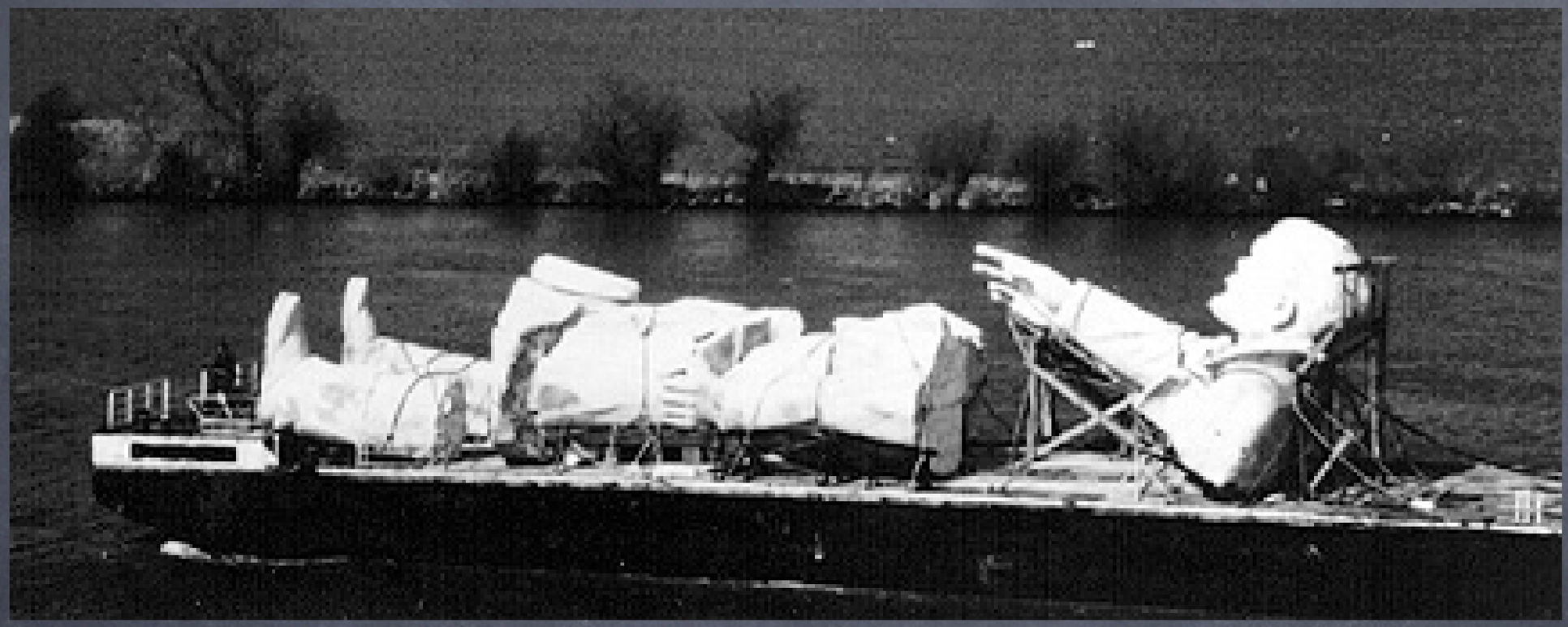
3. Griechische Lettern.

Γιάτι ἴνε ẋὲ γέ μπὲ χιελ̓ χιόφτε σ̈ὲντεροὺαρε ε̈μερι ἴτ· ὄρτε μπρε-
τερία γιότε· οὺ πέφτε οὔρδερι ἴτ, σὶ χούντρε πένετε ντὲ χιελ̓ ὰσ̈τοὺ ὲ δὲ
μπὲ δέ· ε̆πνα νάβετ πούẋεν ὲ σούμε ẋὲ νὰ δούχετε πὲρ φὺστενε· ὲ δὲ

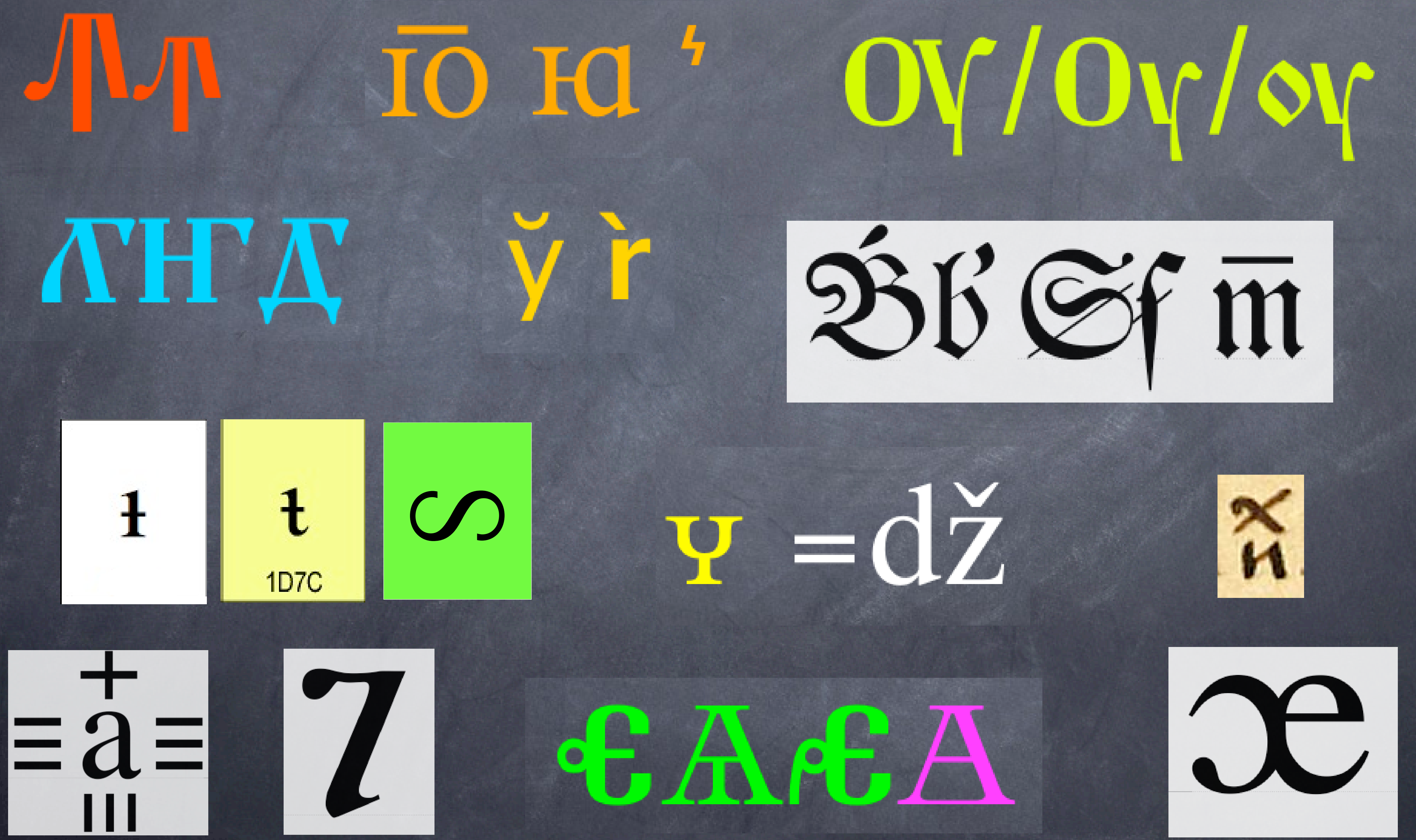
Transscription: *Yati inü tšü ye mbü kiel', kióftü šüntüruarü ümüri it.*
Artü mbretüría yóte, U mbüftü urdüri it, si kundrü mbünetü dü kiel' aštu e de
mbü de. Epna náwet mbutšün e sorme tšü na duχetü per fistünü. E de ndüléna

- Greek for Macedonian…
- Arabic for Byelorussian & Bosnian…

4. Results



Some Results…



Thankъ You!

N.B. The following two papers were the printed outcome of this presentation which was originally given at the corresponding conference:

Sebastian Kempgen: Unicode 4.1 and Slavic Philology - Problems and Perspectives (I). In: A. Miltenova, D. Radoslavova, E. Pancheva (eds.), Computer Applications in Slavic Studies. Proceedings of Azbuky.net. International Conference and Workshop. 24-27 October 2005, Sofia, Bulgaria. Sofia 2006, 131-159.

Sebastian Kempgen: Unicode 4.1 and Slavic Philology - Problems and Perspectives (II). In: T. Berger, J. Raecke, T. Reuther (Hgg.), Slavistische Linguistik 2004/2005, München 2006, 223-248.

Both papers are available online.

ORCID: 0000-002-2534-9423
D-96045 Bamberg, University of Bamberg, Germany
sebastian.kempgen@uni-bamberg.de
https://www.uni-bamberg.de/slavling/personal/prof-em-dr-sebastian-kempgen/