



# SLAVISCHE PHILOLOGIE UND UNICODE 4.1 - PROBLEME UND PERSPEKTIVEN -

Sebastian Kempgen  
Konstanz XXI



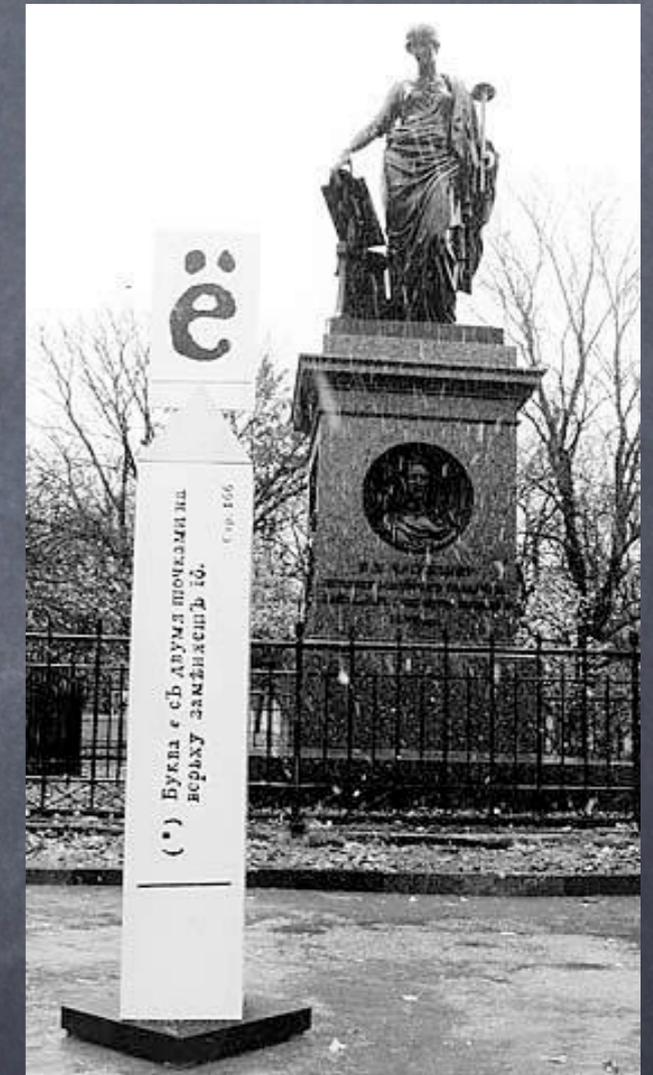
“Kleinschriftsprachen”

“Kleinschriftsprachen”  
'lowercase languages'



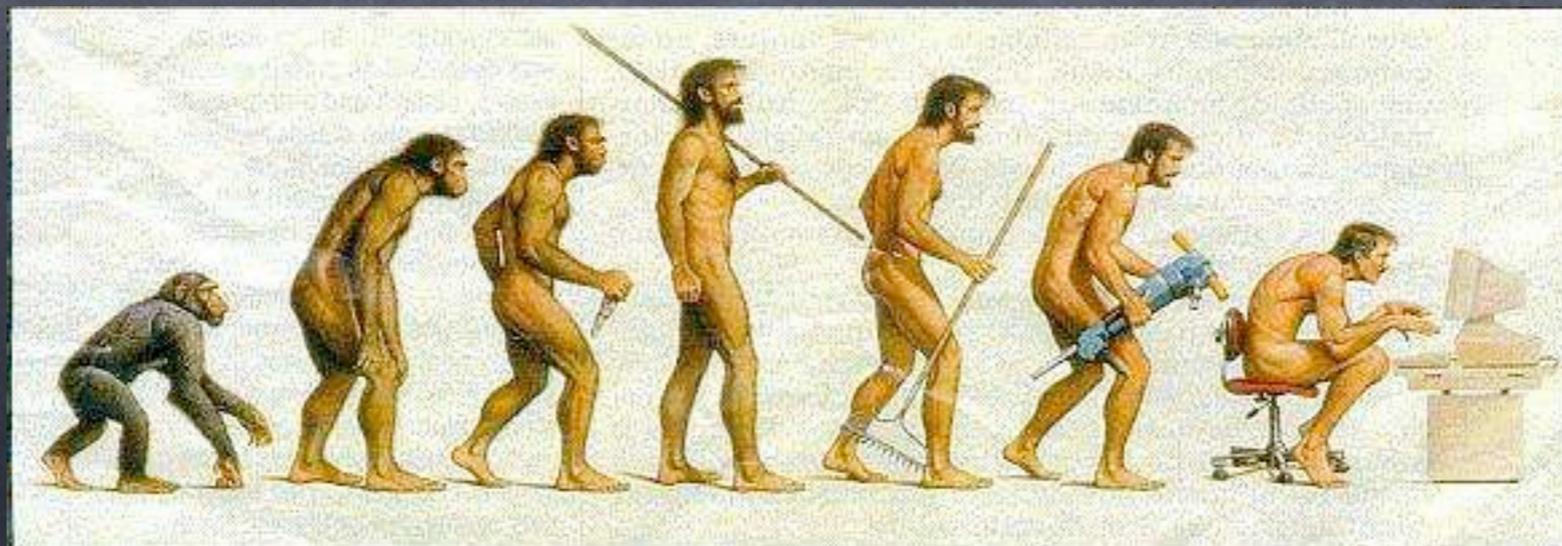
**Alexander Sinin**, russischer Bildhauer, hat eine ungewöhnliche Aufgabe übernommen. Er soll in seiner Heimatgemeinde Ulianowsk ein Denkmal für den Buchstaben „ë“ schaffen - obwohl dieser in offiziellen Dokumenten verboten ist. Der Diphthong, der dem Laut „io“ entspricht, war 1797 von dem aus Ulianowsk stammenden Dichter Nikolai Karamsin eingeführt worden. In der Schriftsprache fallen seine Pünktchen aber häufig weg, in offiziellen Dokumenten darf das „ë“ gar nicht auftauchen, was vielen Familien aus der Region Schwierigkeiten bereitet, weil es in Nachnamen oft vorkommt.

50.8.05  
25



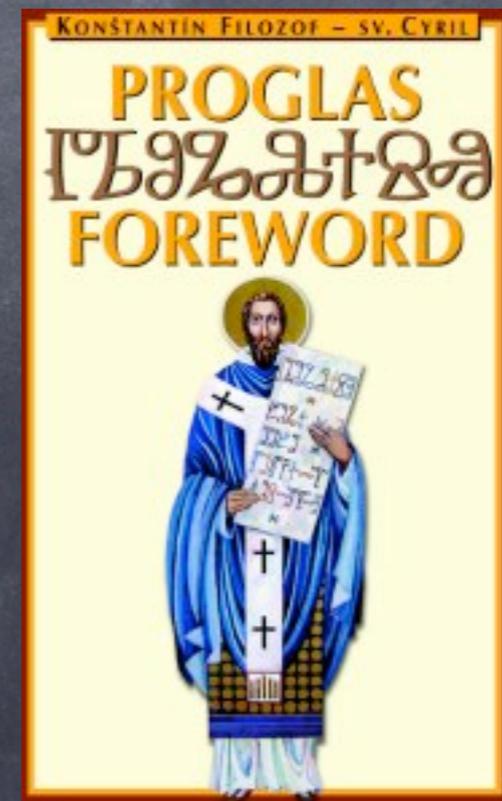
# Overview

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



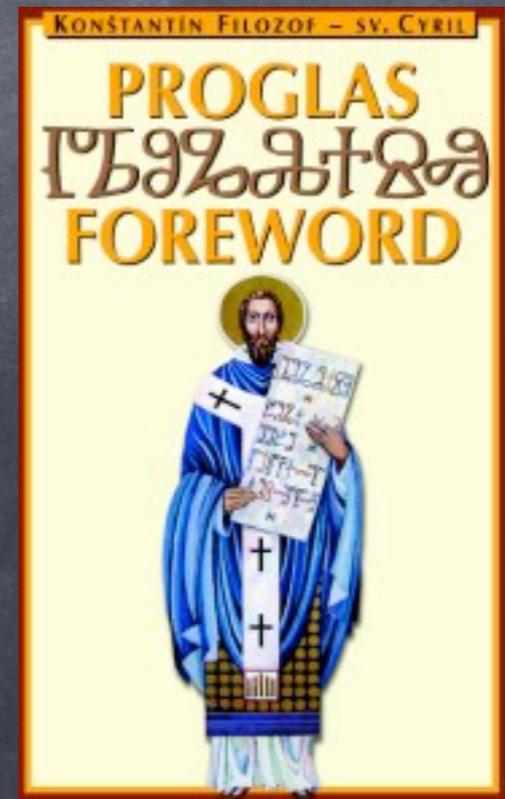
# Thematik des Vortrages

- Linguistik und Informatik
- Datenverarbeitung
- Orthographie, Paläographie
- Phonetik, Phonologie
- Typographie
- Standardisierung



# Bereiche des Vortrags

- Russische hist. Orthographie
- Russische Phonetik
- Polnische hist. Orthographie
- Sorbische hist. Orthographie
- Kroatische Akzentologie
- Bulgarische Phonetik
- Altkirchenslawisch
- Balkanphilologie
- Allgemeine Phonetik
- Mediävistik...



# 1. Introduction

- ASCII:  $2^7$  Zeichen = 128
- 'Code Pages':  $2^8$  Zeichen = 256
- Unicode:  $2^{16}$  Zeichen = 65.536
- Versionen: v. 1.0 1991, v. 4.1 2005
- Meilenstein Version 4.1: Glagolica



# Unicode: Cyrillic

uni040	locyril	Djecyr	Gjecyr	Ecyrill	Dzeczyr	icyrilli	Yicyril	Jecyri	Ljecyr	Njecyr	Tshecy	Kjecyr	uni040	Ushort	Dzhecy
È	Ё	Ђ	Ѓ	Є	Ѕ	І	Ї	Ј	Љ	Њ	Ћ	Ќ	Й	Ў	Ц
Acyril	Becyri	Vecyri	Gecyri	Decyri	Iecyril	Zhecyr	Zecyril	Iicyrill	Iishort	Kacyri	Eicyril	Emocyri	Encyrii	Ocyrill	Pecyri
А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
Ercyrii	Escyrii	Tecyri	Ucyrill	Efcyrii	Khacyr	Tsecyr	Checyr	Shacyr	Shchac	Hardsig	Yericyr	Softsig	Erevert	IUcyrill	IAcyrii
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
acyrill	becyri	vecyri	gecyri	decyri	iecyri	zhecyr	zecyril	iicyrill	iishort	kacyri	eicyril	emocyri	encyrii	locyrill	pecyri
а	б	в	г	д	е	ж	з	и	й	к	л	м	н	о	п
ercyrii	escyrii	tecyri	ucyrill	efcyrii	khacyr	tsecyr	checyr	shacyr	shchac	hardsig	yericyr	softsig	erevert	iucyrill	iacyrii
р	с	т	у	ф	х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
uni045	iocyrii	djecyr	gjecyr	ecyrill	dzeczyr	icyrilli	yicyril	jecyrii	ljecyri	njecyr	tshecy	kjecyr	uni045	ushort	dzhecy
è	ё	ђ	ѓ	є	ѕ	і	ї	ј	љ	њ	ћ	ќ	й	ў	ц
Omega	omega	Yatcyr	yatcyr	Eiotifie	eiotifie	Yuslitt	yuslitt	Yuslitt	yuslitt	Yusbig	yusbig	Yusbig	yusbig	Ksicyr	ksicyri
Ω	ω	Ѡ	ѡ	Є	є	А	а	Ӏ	Ӂ	Ӏ	Ӂ	Ӏ	Ӂ	Ӏ	Ӂ
Psicyr	psicyri	Fitacyr	fitacyr	Izhitsa	izhitsa	Izhitsa	izhitsa	Ukcyri	ukcyrii	Omega	omega	Omega	omega	Otcyrii	lotcyrii
Ψ	ψ	Ѳ	ѳ	Ѵ	ѵ	Ѷ	ѷ	Ѹ	ѹ	Ѻ	ѻ	Ѽ	ѽ	Ѿ	ѿ
Koppac	koppac	thousar	titlocyr	palatal	dasiapr	psilipne	uni048	uni048	uni048	Iishort	Iishort	Semis	semiso	Ercyrii	Iercyrii
Ѕ	ѕ	Ѵ	ѵ	Ѷ	ѷ	Ѹ				Ӏ	Ӂ	Ӏ	Ӂ	Ӏ	Ӂ



# Some 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
- Kompatibilität nicht garantiert
- Benutzung auf eigene Gefahr
- Schon in Gebrauch
- Gebrauch unkoordiniert



# 'Private Area'

Wenn mehrfach in Anspruch genommen, dann ...



- Ist echter Bedarf da für das betreffende Zeichen, und/oder
- Es ist mehr Information notwendig über Grapheme und stilistische Varianten, über Unicode-Grundlagen



# 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.

Translit. for historical  
letters:

Өө: Ĳĳ [ 1E1E, 1E1F ]

Ѳѳ: Ÿÿ ✓ [ 1E8E, 1E8F ]

(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 ´ modifier letter acute accent  
→ 0301 ˆ combining acute accent  
→ 0374 ´ 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 ˆˆ combining double acute accent  
→ 2033 " double prime



(Spacing Modifiers)



# Macedonian Translit.

1E30    **Ǫ**    LATIN CAPITAL LETTER K WITH ACUTE  
          ≡ 004B K 0301 6

1E31    **ǫ**    LATIN SMALL LETTER K WITH ACUTE  
          • Macedonian transliteration  
          ≡ 006B k 0301 6



01F4    **Ǧ**    LATIN CAPITAL LETTER G WITH ACUTE  
          ≡ 0047 G 0301 6

01F5    **ǧ**    LATIN SMALL LETTER G WITH ACUTE  
          • Macedonian and Serbian transliteration  
          ≡ 0067 g 0301 6



(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 źŻ



# Nasal Vowels

Ąą Ęę ĩ ĩ Ūū – Polish orthogr.,  
Lithuanian

Qq – Sami, Old Icelandic

Q̄q̄ – Old Icelandic



ǫ ǣ – not precomposed



# Long / Short Vowels

ā ē ī ō ū ū̄ – long vowels ✓

ă ę ĭ ǫ ŭ \_ – short vowels ✓

ŷ – not precomposed in Latin, available in Cyrillic (for Byelorussian) ✗



# 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 ř ● Latin 1

à è ì ò ù \_ ● Latin 1

â ê î ô û ř ● Extended B

ă ă ĭ ồ ừ ử ● Extended B

ř missing from UC!



# Štokavian Accents

random examples...

S̑rbija  
h̑rvatski

	rising	falling
long	ˀ	ˁ
short	ˁ	ˀ



# Štokavian Accents

## Stokavian accents in Serbian

бр̑нати, бр̑н̑ам *уст.* боронѣть.  
бр̑ндуша *ж.*, бр̑н̑ѣушка *ж.* *бот.* шафр̑ан.  
бр̑нка *ж.* р̑ожа, р̑ожистое воспалѣние.  
бр̑ња *ж.* 1) бѣлое пятн̑о (*на морде живот-*  
*ного*); 2) коз̑а с бѣлым пятн̑ом на головѣ; л̑о-  
шадь с бѣлым пятн̑ом на м̑орде.

...same as in Croatian!



# Štokavian Accents

## Stokavian accents in Serbian

бр̑нати, бр̑н̑ам *уст.* боронѣть.  
бр̑ндуша *ж.*, бр̑н̑ѣушка *ж.* бот. шафр̑ан.  
бр̑нка *ж.* р̑ожа, р̑ожистое воспалѣние.  
бр̑ња *ж.* 1) бѣлое пятн̑о (*на морде живот-  
ного*); 2) коз̑а с бѣлым пятн̑ом на головѣ; л̑о-  
шадь с бѣлым пятн̑ом на м̑орде.

...same as in Croatian!

...but not available in UC



# Štokavian Accents

	rising	falling
long	ˀ	ˁ
short	ˁ	˂˂

Latin: 23 of 24

á é í ó ú ́ • Latin 1  
à è ì ò ù ̀ • Latin 1  
â ê î ô û ˆ • Extended B  
à è ì ò ù ˘ • Extended B

Cyrillic: nothing so far



# Croatian Digraphs

## matching Serbian Cyrillic

uhorn	Upsilon	Vhook	Yhook	yhook	Zstroke	zstroke	Ezh	Ezhreverse	ezhreverse	ezhtail	twostroke	Tonefive	tonefive	glottal	wynn
u	Ϸ	U	Y	y	Z	z	Ʒ	Ʒ	Ʒ	Ʒ	Ʒ	Ʒ	Ʒ	ʈ	ɰ
clickdental	clicklateral	clickalveolar	clickretroflex	Dzcaron	Dzcaron	dzcaron	Lj	Lj	lj	Nj	Nj	nj	Acaron	acaron	Icaron
l	ll	≠	!	DŽ	Dž	dž	LJ	Lj	lj	NJ	Nj	nj	Ǻ	ǻ	Ǫ
icaron	Ocaron	ocaron	Ucaron	ucaron	Udieresis	udieresis	Udieresis	udieresis	Udieresis	udieresis	Udieresis	udieresis	eturned	Adieresis	adieresis
ı	Ǫ	ǫ	Ǫ	ǫ	Ū	ū	Ū	ú	Ū	ǔ	Ū	ù	ə	Ǻ	ā
Adotmacron	adotmacron	AEmacron	aemacron	Gstroke	gstroke	Gcaron	gcaron	Kcaron	kcaron	Ogonek	ogonek	Ogonek	ogonek	Ezhcaron	ezhcaron
Ā	ā	Æ	æ	G	g	Ć	ć	Ǫ	ǫ	Ų	ų	Ų	ų	Ǻ	ž
ıcaron	DZ	Dz	dz	Gacute	gacute	Hwair	Wynn	uni01F8	uni01F9	Aringacute	aringacute	AEacute	aeacute	Ostroke	ostroke
Ĳ	DZ	Dz	dz	Ć	ć	H	ƿ	Ń	ń	Å	å	Æ	æ	Ø	ø
Adblgrave	adblgrave	Ainverted	ainverted	Edblgrave	edblgrave	Einverted	einverted	Idblgrave	idblgrave	Iinverted	iinverted	Odblgrave	odblgrave	Oinverted	oinverted
À	à	Â	â	È	è	Ê	ê	Ì	ì	Î	î	Ò	ò	Ô	ô



(Latin-Extended B)



# Cyrillic 'UK'

Ksicyrillie	ksicyrillie	Psicyrillie	psicyrillie
Ѹ	Ѹ	Ѹ	Ѹ
izhitsadblgrav	Ukcyrillie	ukcyrillie	Omegaroundcy
Ѹ̂	Ѹ	Ѹ	Ѹ
Koppacyrillie	koppacyrillie	thousandseyri	titloeyrilliecn
Ѹ	Ѹ	Ѹ	Ѹ

Ksicyrillie	ksicyrillie	Psicyrillie	psicyrillie
Ѹ	Ѹ	Ѹ	Ѹ
izhitsadblgrav	Ukcyrillie	ukcyrillie	Omegaroundcy
Ѹ̂	ѸѸ	ѸѸ	Ѹ
Koppacyrillie	koppacyrillie	thousandseyri	titloeyrilliecn
Ѹ	Ѹ	Ѹ	Ѹ

two variants:  
ligature or digraph



# Cyrillic OU/ou

Ksicyrilllic	ksicyrilllic	Psicyrilllic	psicyrilllic
Ѣ	ѣ	Ѥ	ѥ
izhitsadbigrav	Ukcyrilllic	ukcyrilllic	Omegaroundoy
Ѧ	Ѧ	Ѧ	Ѧ
Koppacyrilllic	koppacyrilllic	thousandscyri	titlocyrilllicon
Ҁ	ҁ	҂	҃

0468	0469	046A	046B	046C	046D
Ѣ	ѣ	Ѥ	ѥ	Ѧ	ѧ
0476	0477	0478	0479	047A	047B
Ѧ	Ѧ	Ѧ	Ѧ	Ѧ	Ѧ
0484	0485	0486	0487	0488	0489
	Ѧ	Ѧ			

three forms needed

– just as for the Croatian digraphs!

UPPER CASE	MIXEd	lower case
OY	Oy	oy



# Cyrillic OU/ou

Mistake in Unicode 4.1 docs:

0478 Оу **X** 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

ИѦ	Оу	⦿
0468	0478 <b>X</b>	0488
ИѦ	OU	⦿
0469	0479	0489

Ksicyrillic	ksicyrillic	Psicyrillic	psicyrillic
Ѣ	ѣ	Ѳ	ѳ
izhitsadblgrav	Ukcyrillic	ukcyrillic	Omegaroundcy
Ѳ	OU	ou	⦿
Koppacyrillic	koppacyrillic	thousandscyri	titlocyrllicon
Ҁ	ҁ	҂	҃



# Cyrillic OU / ou

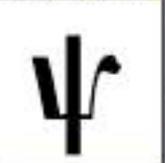
3 forms are needed;

but...

if only 2 are available,

they must be OU/ou,

not Ou/ou, because...

Ksicyrillie	ksicyrillie	Psicyrillie	psicyrillie
			
izhitsadblgrav	Ukcyrillie	ukcyrillie	Omegaroundcy
			
Koppacyrillie	koppacyrillie	thousandseyri	titlocyrillicen
			



# Cyrillic OU / ou

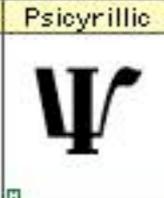
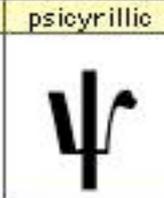
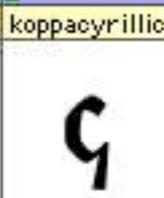
3 forms are needed;

but...

if only 2 are available,

they must be OU/ou,

not Ou/ou, because...

Ksicyrillie	ksicyrillie	Psicyrillie	psicyrillie
			
izhitsadblgrav	Ukcyrillie	ukcyrillie	Omegaroundcy
			
Koppacyrillie	koppacyrillie	thousandscyri	titlocyrillicen
			

this is not acceptable

**✗** ОУГОДИТИ  
ОУГОДИТИ



# Greek OU / ou

strange omission

from a Slavacist's point of view...

no 'ou' ligature in Unicode-Greek!

88



# Russian Phonetics

just added in UC 4.1:

ŷ	Ʒ
1D6B	1D7B
ƚ	ƚ
1D6C	1D7C
đ	Ɔ
1D6D	1D7D

- Ʒ LATIN SMALL CAPITAL LETTER I WITH STROKE
  - used with different meanings by Americanists and Oxford dictionaries
- ƚ LATIN SMALL LETTER IOTA WITH STROKE
  - used by Russianists



# Russian Phonetics

just added in UC 4.1:

ŷ	Ʒ
1D6B	1D7B
ƚ	ƚ
1D6C	1D7C
đ	Ɔ
1D6D	1D7D

- Ʒ LATIN SMALL CAPITAL LETTER I WITH STROKE
  - used with different meanings by Americanists and Oxford dictionaries
- ƚ LATIN SMALL LETTER IOTA WITH STROKE
  - used by Russianists

- 1) wrong
- 2) wrong



# Russian Phonetics

really needed for [ы]:

**і**  
0268



needed and available

**ѣ**



needed but not available

**ѣ**  
1D7C



available but nonsense



# Russian Phonetics

1		2		3
nach API	nach Avanesov	nach API	nach Avanesov	
[ɪ]	[и <sup>е</sup> ]	/e/ /a/ /o/	/e/ /a/ /o/	в лесу́ пяти́, часте́й село́ лесово́д, о лесе́ пятиле́тка, уча́сть мо́рем, перенесена́ би́ты шесты́ дано́ гото́в шестикла́ссник вы́дан, кни́га, пу́ля на́ год, ле́том, мо́рем
[ɪ]	[ь]	/e/ /a/ /o/	/e/ /a/ /o/	
<b>dotless ɪ, not iota!</b>				
[ɪ]	[ы]	/i/	/и/	
[ɪ]	[ы <sup>е</sup> ]	/e/	/e/	
[ʌ]	[ʌ]	/a/ /o/	/a/ /o/	
[ə]	[ъ]	/e/ /a/ /o/	/e/ /a/ /o/	

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<sup>е</sup>л'ѐ<sup>і</sup>ѣт<sup>1</sup> па́рус / лд'инóкѣ<sup>2</sup> /  
ф-тума́н'ь мо́р'ъ / гълубóм //  
штó ѡш':ѣт-он / ф-стран'ѐ длл'óкѣ<sup>і</sup> /  
штó к'ѡнул-он / ф-кря́ѣ рлдно́м //

игра́ѣут во́лны / в'ѣт'ѣр с'в'ѡш':ѣт /  
и-ма́ч'тъ / гн'óтѣцъ и-скрып'ѣт<sup>3</sup> /  
увѣ<sup>і</sup> / он-ш':ѡс'т'ѣѣ<sup>4</sup> н'ь-ѡш':ѣт /  
и-н'ь-лт-ш':ѡс'т'ѣѣ<sup>4</sup> / б'и<sup>е</sup>жѣт /

плд-н'ѣм стру́ѣ / с'в'и<sup>е</sup>тл'ѐ<sup>і</sup> лзѣѣр'и /  
налд-н'ѣм / лу́ч'-со́нѣцъ зѣлтóѣ<sup>і</sup> /  
л-он м'и<sup>е</sup>т'ѣжнѣѣ<sup>5</sup> / прóс'ѣт бѣѣр'и /  
клг-бѣѣтѣ<sup>6</sup> в-бѣѣр'ѣх ѣѣс'т' плкóѣ<sup>і</sup> ///

Avanesov  
1972

Cyrillic +  
Latin +  
Greek +  
Diacritics



# Russian Phonetics

Cyrillic	абвгде...	✓
Latin/IPA	h i̇ j ə л ѝ	✗
Greek	γ α	✓
Diacritics	^ / \ • •• ...	✗

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



# Russian Phonetics

Additional character 'Nasal':  $\text{ɲ}$  [223D]

Самостоятельным, фонематическим признаком является „носовость вообще“ без локализации его образования. В словофонематической транскрипции носовую согласную фонему слабости по признаку дентальности-лабиальности обозначим знаком  $\text{ɲ}$ . 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  $\text{ɲ}$  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 $\text{ɲ}_1$ т $_2$ |, |ба $\text{ɲ}_1$ т $_2$ |, |ба́ $\text{ɲ}_1$ да|, |кама́ $\text{ɲ}_1$ д $_2$ а|, |в'и́ $\text{ɲ}_1$ т'-ик $_2$ |, |ба́ $\text{ɲ}_1$ т'ик $_2$ |; |ра́ $\text{ɲ}_1$ па|, |ла́ $\text{ɲ}_1$ па|, |а́ $\text{ɲ}_1$ ба́р|, |бо́ $\text{ɲ}_1$ ба|, |ба́ $\text{ɲ}_1$ б'и́т' $_2$ |.

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



# Russian Phonetics

Just accepted ('Pipeline'):

1DC4..1DCA	7	COMBINING MACRON-ACUTE COMBINING GRAVE-MACRON COMBINING MACRON-GRAVE COMBINING ACUTE-MACRON COMBINING GRAVE-ACUTE-GRAVE COMBINING ACUTE-GRAVE-ACUTE COMBINING LATIN SMALL LETTER R BELOW	2004-Aug-11 Accepted	2005-Jan-28 Stage 5
------------	---	--	-------------------------	------------------------

◌́ etc. ✓



# Slavic Phonetics

Also needed...



– Russian:

ˆ ѝэѐ

˙˙ аоѳ

– Proto-slavic/  
Indo-European

˘

'long or short'

ˆ

'long w. circumflex'



# Bulgarian Phonetics

## Reviving OCS characters:

- # За означаване на дълга пауза в края на изречението:  
и дъждове #.
- Ѹ За означаване на африкат дж: [Ѹоп], [ѸуѸè], [Ѹас].
- s За означаване на африкат дз: [sънкам], [сифт].
- l За означаване на изговора на средно [л] пред гласна  
[lèсно].

s = dz ✓

Ѹ = dž ✗

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



# Sorbian Orthography

Some Sorbian Characters:

ń p r ě ě ž ł m w



Ǻ ǻ ǣ ǿ Ǿ ǿ



(Latin Extd.-A, Extd. Additional)



# Sorbian Orthography

## I. Lautlehre

### § 1. Buchstaben

Das Niedersorbische wird heute ausschließlich mit lateinischen Buchstaben geschrieben. Da bis 1933, wo der Nazismus jede niedersorbische Kulturarbeit erstickte, neben der erwähnten Schreibweise auch mit deutschen Buchstaben, letzteres sogar häufiger, geschrieben wurde, sei zum Verständnis der älteren Schriften auch diese Schreibweise mit angegeben.

a) Lateinische Schreibweise:	a	b	ǰ	c	č	d	dž	dź	ž	e	ě	f	f'	g	h												
b) deutsche Schreibweise:	a	b	ǰ	c	č	d	dž	dź	ž	e	ě (ë)	f	f'	g	h												
	ch	i	j	k	ł	l	m	m̃	n	ñ	o	p	p̃	r	r̃	(ř)	s	š	t	ś	ć	u	w	w̃	y	z	ž
	ch	i	j	k	ł	l	m	m̃	n	ñ	o	p	p̃	r	r̃	(ř)	s	š	t	ś	ć	u	w	w̃	y	z	ž

Die Benennung der Buchstaben entspricht dem Deutschen; dazu merke: ć, dž, ś, ź nennt man ćej, dzej, śej, źej; č, ž, ě = čet, žet, ět, š = eš; ł = eł; ch, dž = cha, dža<sup>1</sup>.



# Sorbian Orthography

- z, ʃ weiches s (lesen) za – ʃa für (**S**ache), zły – ʃły böse.
- s, ʃ hartes s, ʃ (Grü**ß**e, mi**ß**): se – ʃe sich (Erbse), gus – gu**ß** die Gans (der **G**u**ß**).
- c, ʒ wie dt. z: cas – ʒa**ß** die Zeit, cera – ʒera Linie (**z**erren).
- ź, ʒ ganz weiches dt. sch mit j verbunden: źiwy – ʒiwy wild.
- ś, ʃch dt. sch mit j verbunden: śele – ʃchele Kalb, śpa – ʃchpa Stube.
- ć, cʒ dt. tsch mit j verbunden = tś: ścina – ʃcʒina Rohr, ścaniś – ʃcʒaniśch schillern.
- dź, dʒ = d + ź: zdźaržaś – ʃdʒarʒaśch erhalten.
- ž, ʒ volles sanftes sch, wie im Worte journal: žaba – ʒaba der Frosch, żywy – ʒiwy lebendig.
- š, ʃch härter als dt. sch (Pas**ch**a), šeść – ʃchescʒ sechs, šnarl – ʃchnarl der Goldammer (**s**chnarren), běšćo – běšchcʒo ihr waret.
- č, cʒ volles tsch (K**u**t**s**cher) = tš: lažčej – laʒcʒej leichter.
- dž, dʒ = d + ž: ldžej – ldʒej leichter.
- ř, ʃch = sch, steht in manchen älteren Schriften nach k, p, t statt heutigem š: křej – kšej, třawa – tšawa.



# Sorbian Orthography

18

TABELLARISCHE UEBERSICHT DER NS. ALPHABETE.

Fabricius 1706	Fryco 1796	Zwahr 1847	Tešnar- Šwjela	Časopis M. S.	Mein Alphabet	Ober- sorbisch	Alt- slovenisch
p	p	p	p	p	p	p	p
r	r z	r	r	r	r	r	r
—	sch — fch	—	—	ṛ	—	ṛ	—
ß (ff)	ß	ss	ß	s	s	s	s
fch	ßch	sch	fch	ṣ	ṣ	ṣ	ṣ
fch	fch	schj	fch	ṣ	ṣ	—	—
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



# Polish Orthography

breue	longum	groffum	molle		
<i>a</i>	<i>aa</i>	<i>b</i>	<i>b</i>		
c has quinque differencias in Polonico idiomate habet					
	<i>c</i>	<i>k</i>	<i>g</i>	<i>cz</i>	<i>ch</i>
per se	molle	breue	longum	durum	molle
<i>d</i>	<i>dz</i>	<i>e</i>	<i>ee</i>	<i>ff</i>	<i>f</i>
per se	inproprie	breue	longum	durum	molle
<i>g</i>	<i>g<sup>87</sup></i>	<i>i</i>	<i>ij</i>	<i>l</i>	<i>l</i>
groffum	molle	groffum	molle	breue	longum
<i>m</i>	<i>m</i>	<i>n</i>	<i>n</i>	<i>o</i>	<i>oo</i>
durum	molle	per se	per se		
<i>p</i>	<i>p</i>	<i>q</i>	<i>R r</i>		
S in lingua Polonorum istas in se sex differencias habet					
	<i>f</i>	<i>s  </i>	<i>ffz</i>	<i>sch</i>	<i>sz</i> <i>zz</i>
		breue	longum		
<i>t</i>	<i>u</i>		<i>uu</i>	cum perdit vim vocalitatis,	
has tres habet differencias, que differencie patent in abecedario Polonorum, scilicet adaam bił etc.					
	<i>v</i>	<i>v</i>	<i>w</i>	<i>x</i>	<i>y   </i>



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





# Polish Orthography

¶ Alphabetū Polonicū.  
A ą á vel ǫ ȳ. b b. c ć ċ. d đ. e ē. f. g. h. i. i. k. l l'. m m̄. n ñ. vel sic. m̄. ñ.  
o ó. p p̄. q. r r̄. s ś. t. v ũ. w. x. y. z ź.

[Następuje teraz] Alfabet polski

A ą á lub ǫ ȳ. b b̄. c ć ċ. d đ đ. e ē. f g ḡ. h i ī. k k̄. l l' m m̄. n ñ. lub tak m̄  
ñ. o ó. p p̄. q. r r̄. s ś. t v ü. w w̄. x x̄. y z ź ź.

St. Zaborowski, Orthographia seu modus recte scribendi  
et legendi polonicum idioma quam utilitissimus.  
Cracoviae 1514/1518.



# Polish Orthography

A Á a b **B** c ċ ċ d ḋ dz ḋz e é e f g h i ü j ŷ y i l  
P n̄ m̄ n̄ ñ o ó p p̄ r **r̄** t̄ z s̄ s̄ ī s̄ Ń z̄ z̄ z̄ t̄ ū ū w̄ w̄ x̄.



A Á a b **B** c ċ ċ d ḋ dz ḋz e é e f g h i ü j ŷ y i l  
P n̄ m̄ n̄ ñ o ó p p̄ r **r̄** t̄ z s̄ s̄ ī s̄ Ń z̄ z̄ z̄ t̄ ū ū w̄ w̄ x̄.

St. Murzynowski, Orthographia Polska.  
Königsberg 1551.



# Polish Orthography

## “r rotunda”

Um doppeldeutige Buchstabenverbindungen zu vermeiden, hat Murzynowski das Zeichen **7** eingeführt, da er die Verbindung  $rz = r$  von  $r+z/r+ż$  unterscheiden wollte in Beispielen wie *mierzy/mier-zi*, *skarżę/skarżę*. Der Buchstabe **7** wurde sehr häufig auch vom Schreiber des *Puławer Psalters* verwendet.

St. Murzynowski, Orthographia Polska.  
Königsberg 1551.



# Polish Orthography

OBIĘCĄDŁO WEDŁE ORTHOGRA-  
phiiéy Páná Kochánowskiégo ták idzie.

A a ą **b b̄** c ć ę d d̄ d̄̄ d̄̄̄ é è ę f g h d i j k **l** m m̄ n ñ o  
ò p p̄ q **r r̄** s̄ s̄̄ s̄̄̄ t u w n̄ x y **z z̄ z̄̄.**



J. Januszowski, Nowy charakter polski i  
Orthographia polska  
(Jan Kochanowski, Łukasz Górnicki).  
Cracoviae 1594



# Czech Orthography

Latin, Medieval German etc.

v = u = [u]



# Czech Orthography

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

ů

údolj “valley”

x

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



# Czech Orthography

Needed for Czech: **ů** ‘údolj’

Needed for German: **ü** ‘führet’



# 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)**



# Hist. 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!



# Hist. 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

Ѡ ≈ Щ

etc.



# OCS, Old Russian

functional identity

Л ≈ Љ

Н ≈ Њ

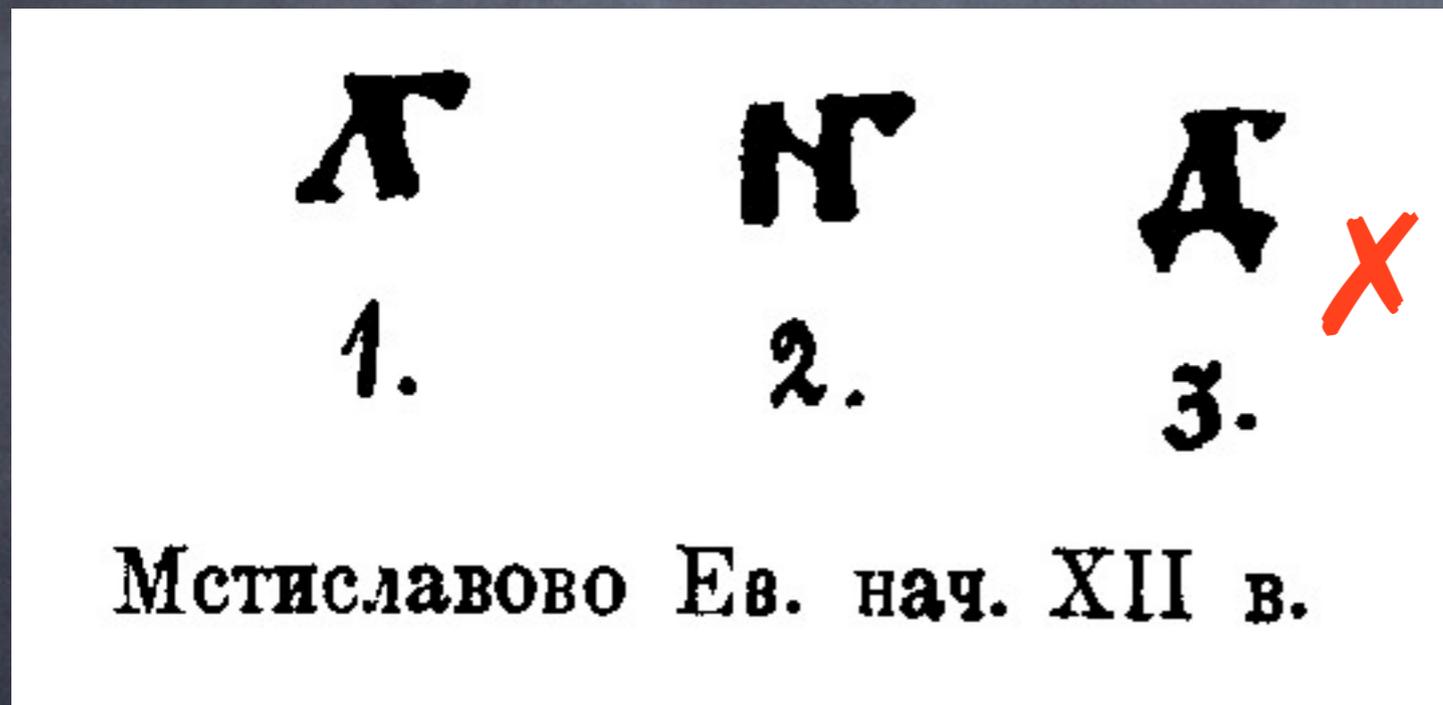
Г̣ ≈ Ѓ

К̣ ≈ Ќ

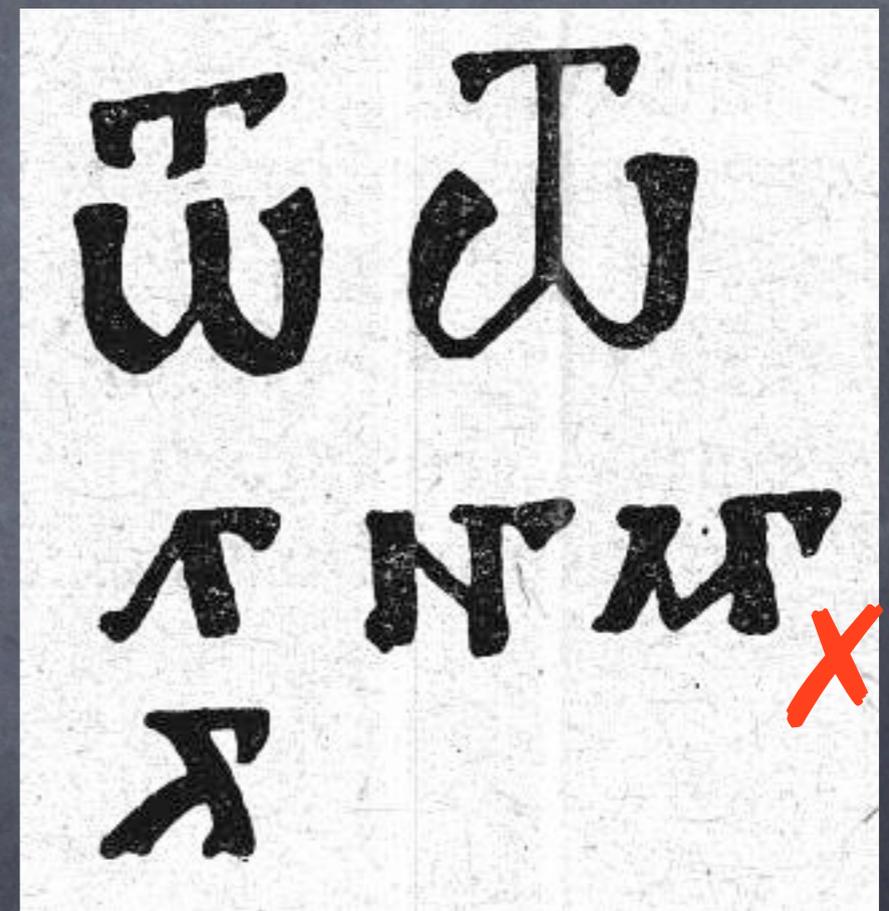


# OCS, Old Russian

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

first problem...

Л ≈ Љ!

Н ≈ Њ!

Д ≈ Ъ?

(because Ъ ≈ Ѓ)



# OCS, Old Russian

functional identity    no correspondence

Л ≈ Љ

Λ ≈ ?

Н ≈ Њ

Д ≈ Ђ

Ṗ ≈ ?

Г ≈ Ґ

Ḡ ≈ ?

К ≈ Ҁ

Ḳ ≈ ?



# OCS, Old Russian

best-known problem...

Ѡ ≈ Я ?

Ѡ ≈ А ?



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



# OCS, Old Russian

best-known problem...

Ѧ ≈ Я ?

Ѧ ≈ А ?



best solution: Ѧ Ѧ separate glyphs!



# Early Modern Russian

	[jo]
18th century (AG-1802)	îô
Karamzin 1797	ë

x

✓

variation: îô, jô, ñô, îô...



# Early Modern Russian

wrong: iô, iô, ôi

acceptable:  $\bar{i}o$   $i\hat{o}$   $i\bar{o}$  x

§ 53. Къ надспрочнымъ знакамъ относятся такъ же:

1) *Краткая* (̣) употребляемая надъ буквою *й* для означенія сокращенія оныя въ произношеніи, на пр: *рой̣*, *безчисленный̣*, и проч.

2) *Слитная* (^) поставляется надъ буквами *io* для означенія слитности въ произношеніи, на пр: *сині̂o*, *Маі̂оръ* и проч.



# Early Modern Russian

	[jo]	[ja]	
'Old Form'	Ѳѳ	Ѵѵ	✗
'Modern Glyph'	ë	я	

Unicode: either both or none -  
too important for private area!



# Early Modern Russian

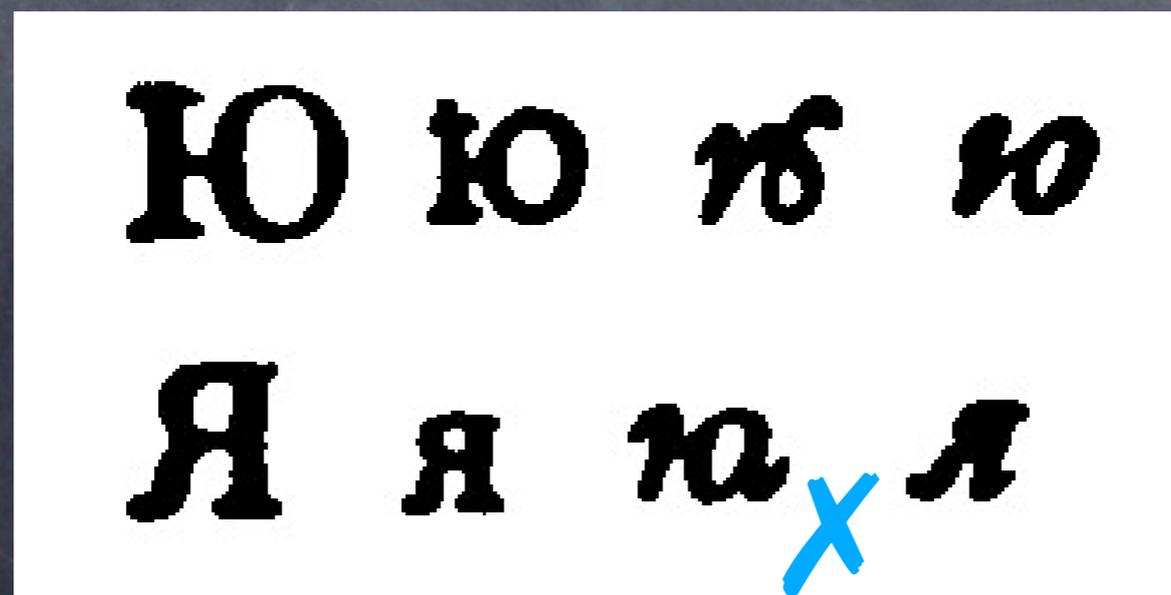
100 years after Peter I removed it from the Russian alphabet...



# Early Modern Russian

100 years after Peter I removed it from the Russian alphabet...

Ю considered cursive form of Я like Т – *т*, Д – *д*, etc.



Россійская  
Грамматика,  
соч. Имп.  
Росс. Акад.,  
СПб. 1802, 3



# Cursive Cyrillic

б = δ

б = δ

Г = Г

г = ġ

Д = Д

д = g

П = П

п = ū

Т = Т

т = ū

Serbian & Macedonian  
*cursive* variant forms.

Russian norms on the left.  
Serbian and Macedonian  
norms on the right.

Note that only one variant,  
the lowercase /be/ occurs  
in both roman and italic styles.  
The other variants occur  
only in the italic style.

Standard  
solution needed  
for cursive  
variants



# Old Russian

Jotated Jat'

ⱭⱮ > ѣ ✕

Separate glyphs or variants of jat?



# Old Russian

Jotated Jat'

Ѣѣ > ѣ ×

Separate glyphs or variants of jat?

Ѧ-ѦѦ, ѧ-ѧѧ, Ѩ-ѨѨ, ѩ-ѩѩ



# Old Russian

Not encoded yet:

Ѧ Ѧ Ѧ Ѧ

Ѧ Ѧ Ѧ Ѧ Ѧ



# Old Ukrainian

ЧТО ЁСТЬ ГРАММАТИКА:

Ёсть и́звѣстное ху́дожество бѣгшѣ  
глаго́лнѣ писати оучащеѣ;

Колѣкш ѡ́сть частѣи грамматики:

Четы́ри: { Ордографѣа:  
                  { Ётѣмолѣа:  
                  { Синтаксѣс:  
                  { Просѡдѣа:



# Old Ukrainian

ЧТО́ ъСТЬ ПИСМА́ ;

ѢСТЬ РЕЧЕНІА ЧАСТЬ НЕРАДѢЛНАЯ .

КОЛІКѠ ѢСТЬ ПИСМЕ́ СЛАВЯ́НѠ ;

ЧЕТЫРДЕСА́ :

а б в г нли у д е ж з з нли з , н і  
к л м н о п р с т оу нли Ѹ , ф х  
ѡ ц ч ш щ ѡ ы в ѣ є нли іє  
ю нли ж , я нли а , ѡ нли о , џ џ а ѣ .



# Old Ukrainian



# Old Ukrainian



А, а	<i>А а</i>		a	a
Б, б	<i>Б б</i>	<i>б̄</i>	be	b
В, в	<i>В в</i>	<i>в̄</i>	we	v, u
Г, г	<i>Г г</i>		ha	h
Ґ, ґ	<i>Ґ ґ</i>		ge	g
Д, д	<i>Д д</i>	<i>д̄ д̄</i>	de	d
Е, е	<i>Е е</i>		e	e

Ja. Rudnyc'kyj, Lehrbuch der ukrainischen Sprache.  
3. Auflage, Leipzig 943, 2.



# Old Ukrainian

Что́ есть́ писма́ ;

Есть́ рече́нїа часть́ нера́дѣлая .

Колѣкѡ́ есть́ писме́ Славя́нѡ́ ;

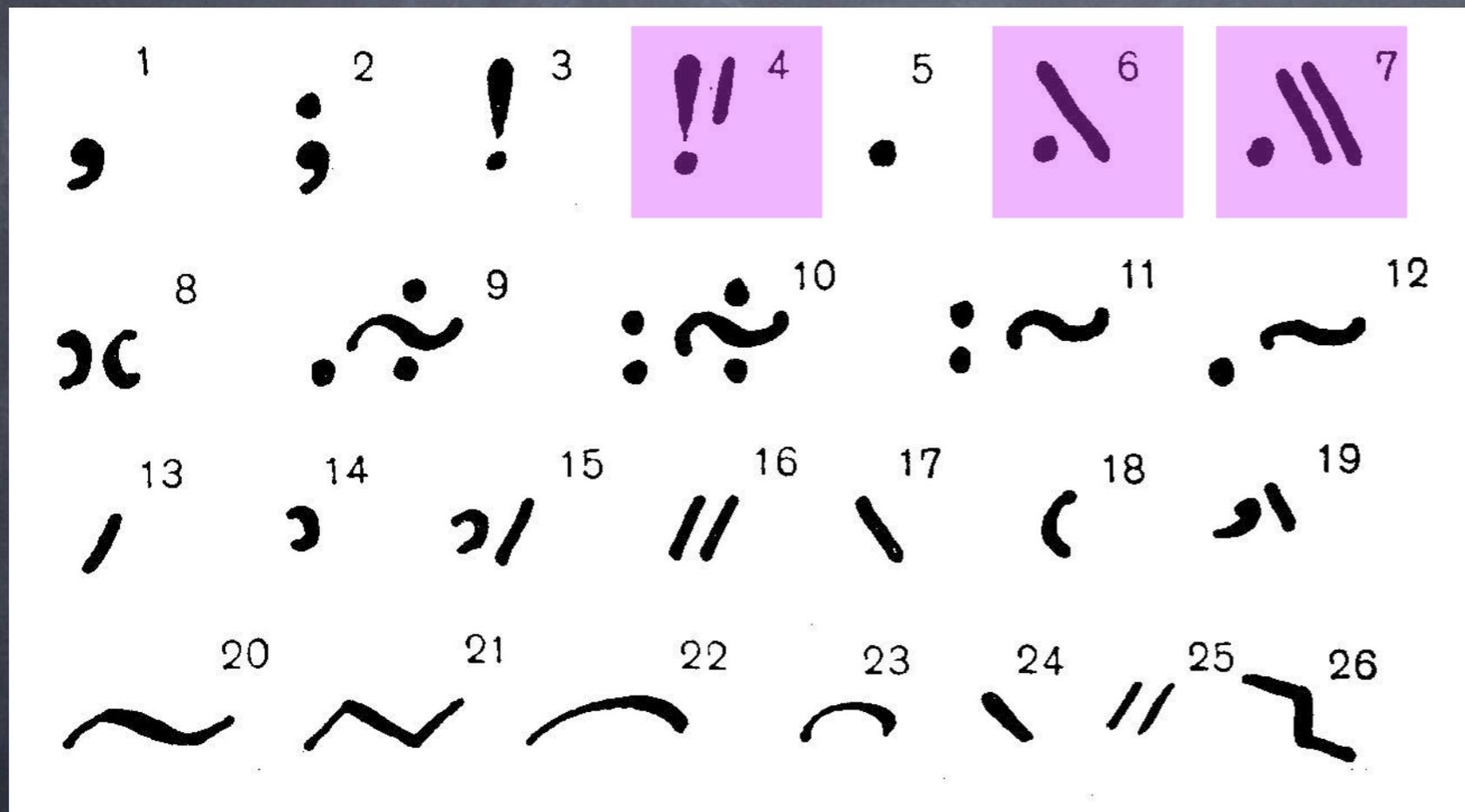
Четырде́ся :

а б в г илї ѓ д е ж з з илї з , н і  
к л м н о п р с т оу илї ѱ , ф х  
ѡ ц ч ш щ љ ы в ѣ є илї іе  
ю илї ж , я илї а , ѡ илї о , ѱ Ѳ а ѱ .



# Old Russian

## znaki prepinanija 1



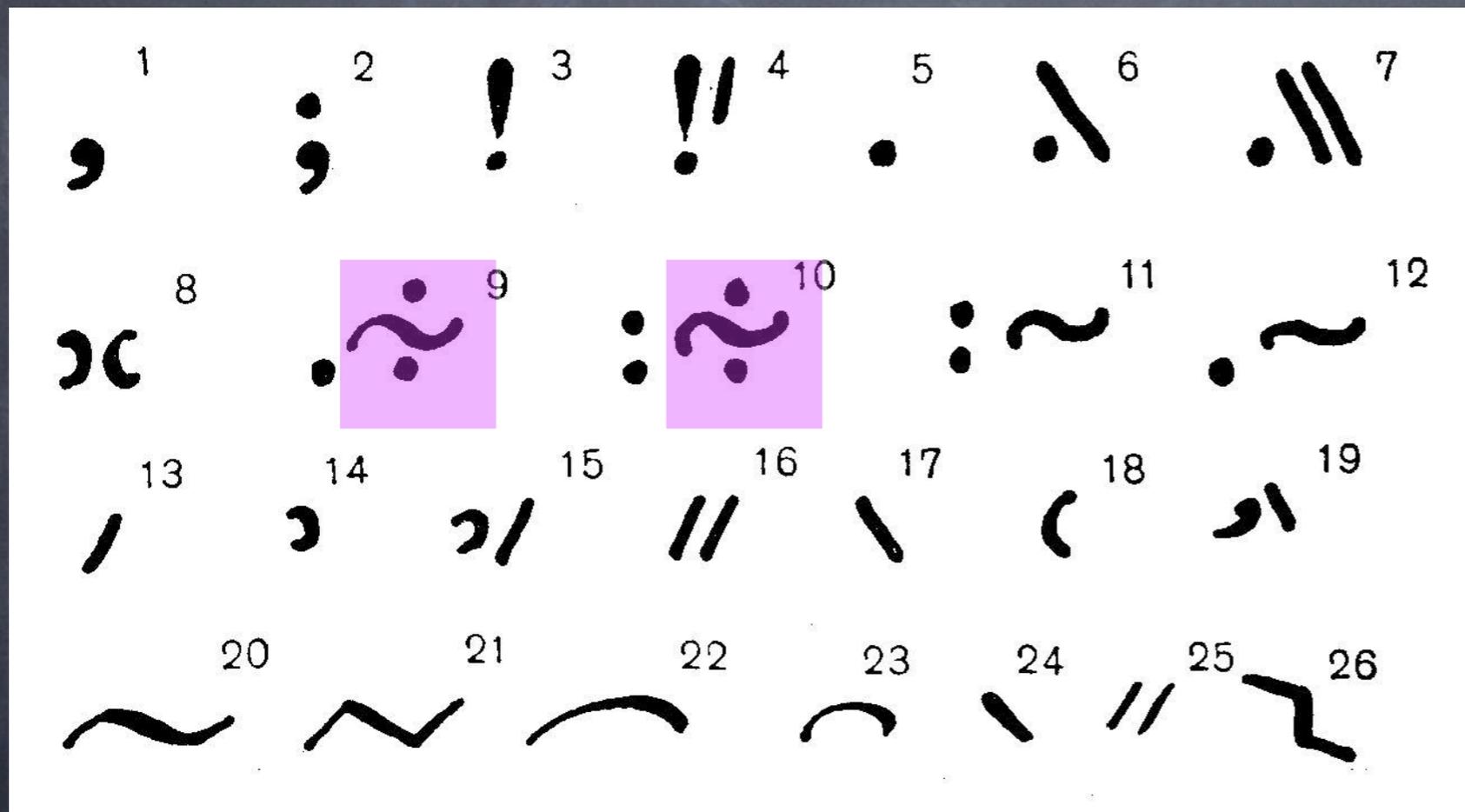
- 1 zapjataja
- 2 podstolija
- 3 pribyl'ca
- 4 pribavlenaja
- 5 točka
- 6 položitel'naja
- 7 kendema

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



# Old Russian

## znaki prepinanija 2



8 kavyka

9 slogija

10 statija

11 stišica

12 otrikal'

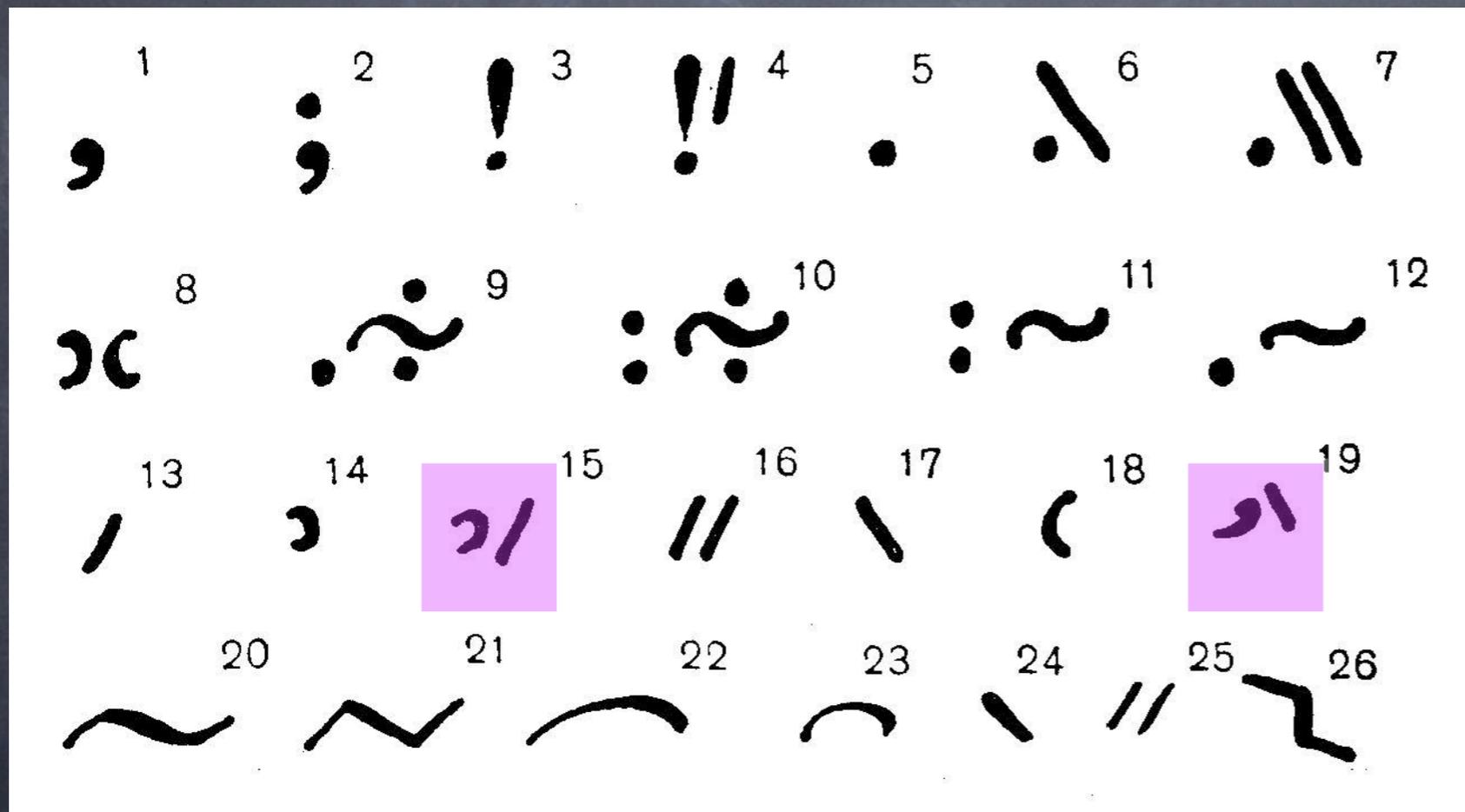


Л.В. Черепнин, Русская палеография,  
Москва 1956, 374-376.



# Old Russian

udarenija i pridyxanija



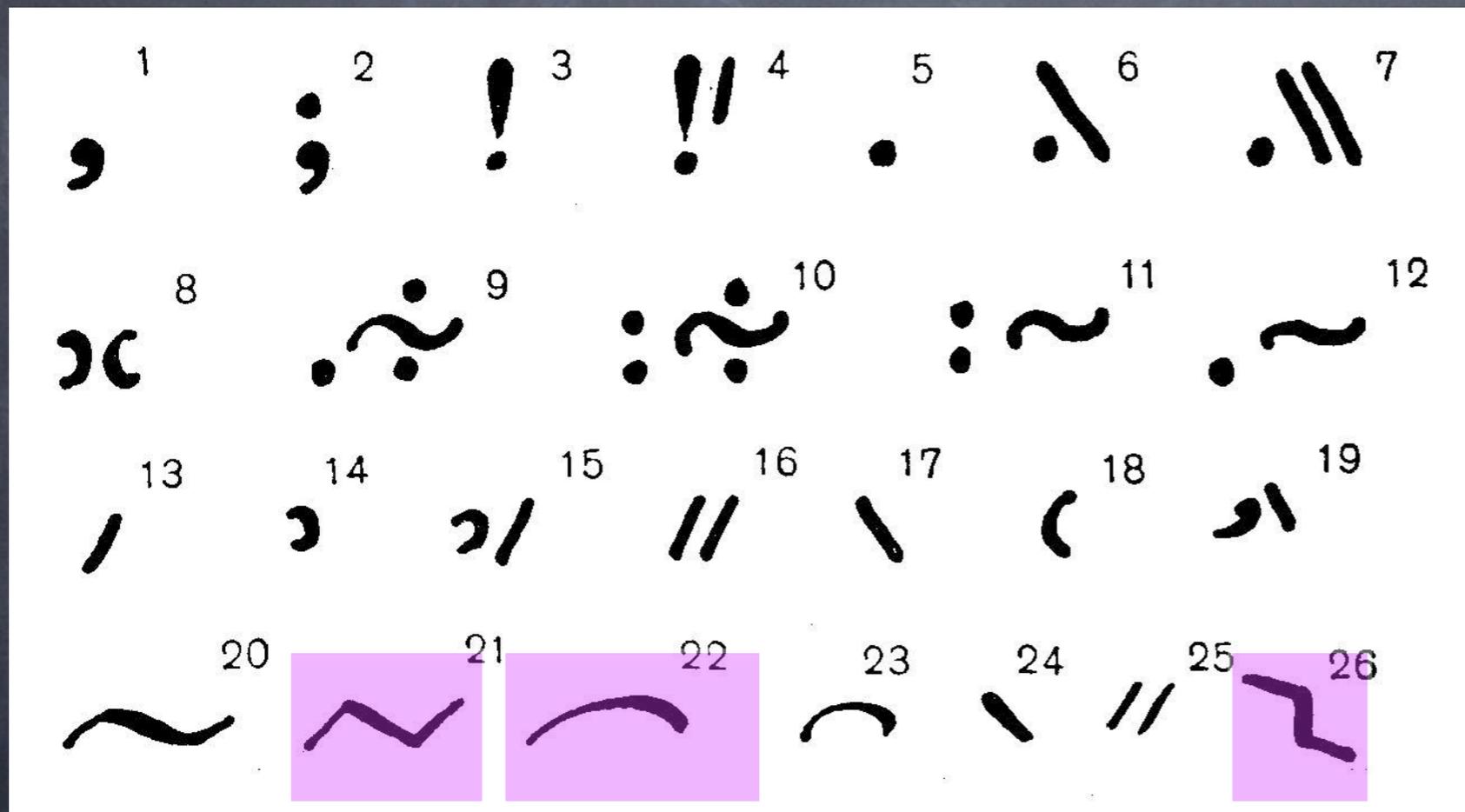
- 13 oksija
- 14 zvatelco/psili
- 15 iso/isso
- 16 okovavy
- 17 varija
- 18 dasija
- 19 apostrof

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

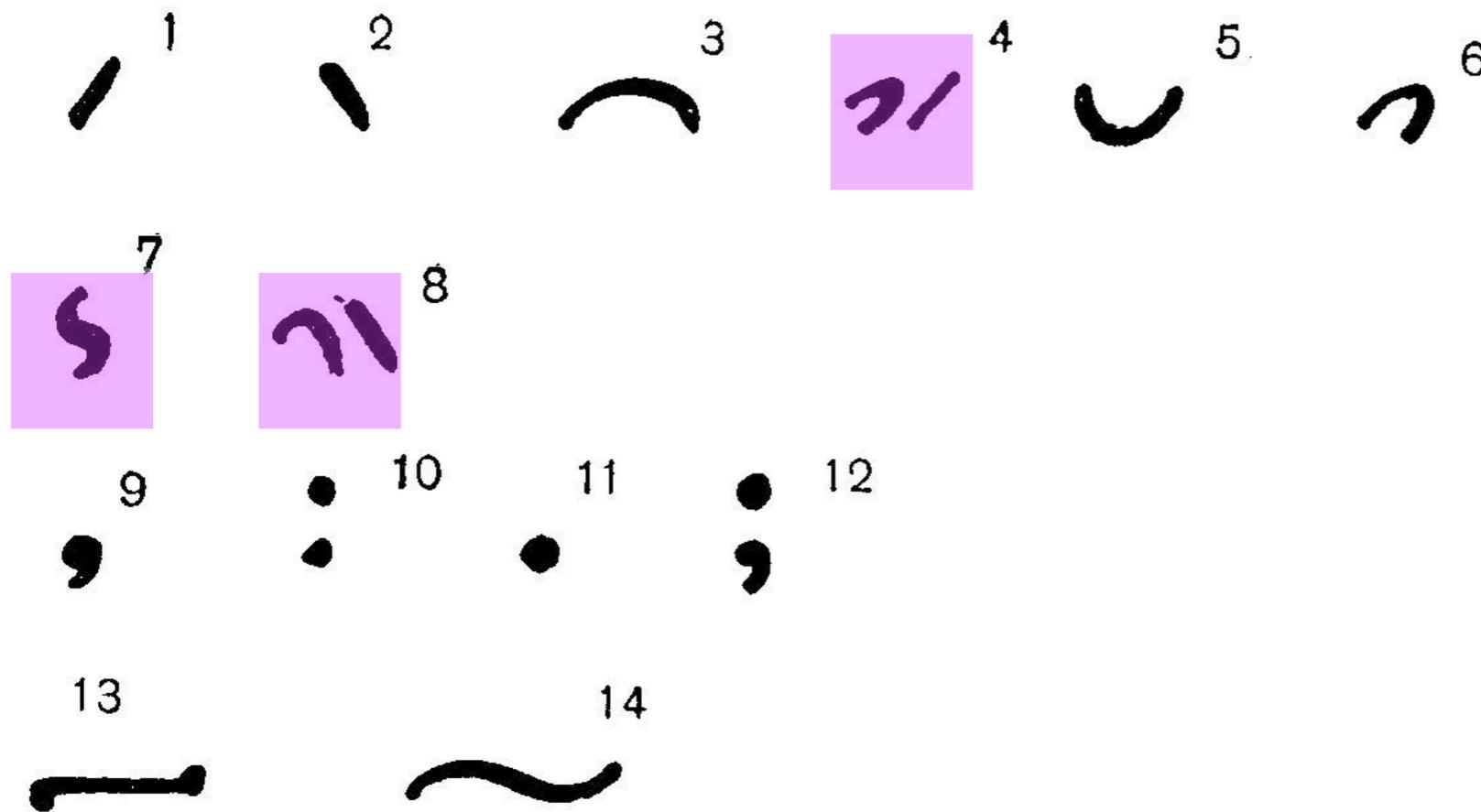
26 erok, ertica

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



# Old Russian

## Smotrickij 1618: Grammatika



1 oksija/ostraja  
2 varija/tjažkaja  
3 perispomena/  
oblečenaja

4 iso/dolgaja

5 kratkaja

6 psilija/gustaja

7 erik, paerk

8 apostrof

Л.В. Черепнин, Русская палеография,  
Москва 1956, 377.



# Old Russian

problem: paerok

1. знак ' или  $\overset{\frown}{\text{~}}$ : мор'е, к'итъ, вѣпль.
2. знак ' или ' : м'ного (мѣного), д'ва (дѣва),  
X в'се (вѣсе), ч'то (чѣто).
3. титло  $\overset{\sim}{\text{~}}$ ,  $\overset{\frown}{\text{~}}$ ,  $\overset{\text{~}}{\text{~}}$ ,  $\overset{\sim}{\text{~}}$ : снѣ (сынѣ), вѣка (владѣка),  
ісѣ (исоусѣ), глѣ (глава),  
мѣць (мѣсѣць).



# Old Russian

problem: paerok

2. знак ' или ' : М'НОГО (МЪНОГО), Д'ВА (ДЪВА),  
X В'СЕ (ВЪСЕ), Ч'ТО (ЧЪТО).

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



# Old Russian

problem: paerok

2. знак ' или ´ : М'НОГО (МЪНОГО), Д'ВА (ДЪВА),  
X В'СЕ (ВЪСЕ), Ч'ТО (ЧЪТО).

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

≠ UC+033E "comb. vertical tilde": ´



# Old Russian

part of the solution

02BC ’ 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'	ĭō	ĭа	⚡
'Modern Glyph'	ë	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
ī	īī	īīī	iv	v	vi	vii	viii	ix	x	xi	xii	l	c	d	m

Roman Numerals: Everything present 



# Old Church Slavonic

СЛОВ'ЯНСЬКІ ЦИФРИ

·ā·	·b̄·	·ḡ·	·d̄·	·ē	·s̄	·z̄	·h̄	·ī	·t̄
1	2	3	4	5	6	7	8	9	10
āī	b̄ī	ḡī	d̄ī	ēī	s̄ī	z̄ī	h̄ī	ī	ī
11	12	13	14	15	16	17	18	19	
k̄	l̄	m̄	n̄	z̄	ō	p̄	č		
20	30	40	50	60	70	80	90		
ṛ	ṛ	ṛ	ṛ	φ	χ	ψ	ω	ц	
100	200	300	400	500	600	700	800	900	
*ā	*b̄	*ḡ	*d̄	*ē					
1000	2000	3000	4000	5000					

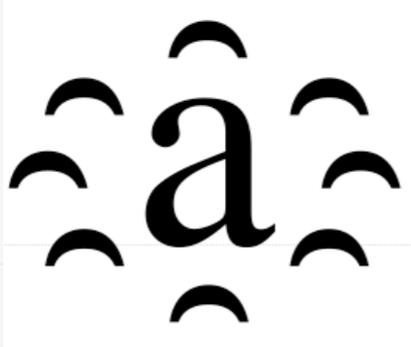
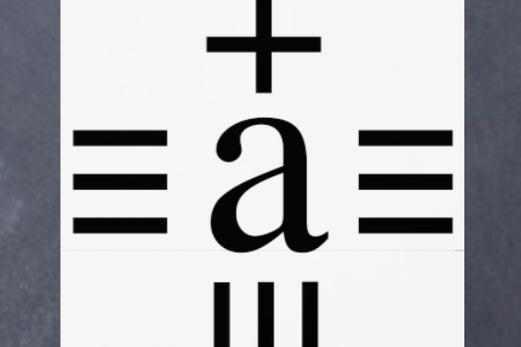
Roman Numerals: Everything present X

Slavic Numerals: Nothing present (≠) X



# 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.

Ɱ ⱮⱮ ⱮⱮⱮ



и ѱ і і



# Glagolitic Translit.

Ɀ Ɱ Ⱳ



и ü ï |



# Glagolitic Translit.

## I. Schrift und Aussprache.

9

§ 6. Umschreibung des glagolitischen Alphabets durch das kyrillische. Bei Herausgabe glagolitischer Denkmäler in kyrillischer Umschrift werden wiedergegeben: **Ѡ durch ѡ** (ein neu erfundenes Zeichen), daher auch ѠѠ durch ѡѡ; ѡ durch ѡ, daher auch Ѡѡ durch ѡѡ; Ѡ durch Ѡ, daher auch ѠѠ durch ѠѠ; Ѡ durch Ѡ, ein aus der altserbisch-kyrillischen Schrift entlehntes, im altbulg. Alphabet nicht gebrauchtes Zeichen; Ѡ, ѡ in den beiden Bedeutungen (= kyrillisch *ě* und = kyrill. *ja á*), durch Ѡ; im übrigen durch die Buchstaben, wie sie in den Alphabeten § 1 einander gegenüberstehen.



# Glagolitic Translit.

## Samples: Cyrillic Iota

ЖАШЕ ЕГО ПРѢДАТИ. АШТЕ НЕ БИ САМЪ ХОТѢЛЪ.  
X И ТОГО НЕ МОЖАШЕ ЗЪРѢТИ. ЕГО <sup>21</sup> И ГЛА-  
ДАТИ. ИБО СВѢТІЛЬНИКОМЪ СЖШТЕМЪ.  
X И СВѢШТАМЪ ТОЛКАМЪ. СЕ БО <sup>22</sup> И ПРИ-  
5 ЕВАНѢЛИСТЪ РЕЧЕ. ꙖКО СВѢТІЛЬНИКЪ  
СВѢШТЪ НОШАХЪ. И ТАКО ЕГО <sup>27</sup> ИЗИДЕ ЖЕ ИСЪ  
САТЪ. И НЮДА СТОѢШЕ СЪ НИМИ. ТЪ РЕКЪ

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

Jagić, Edition of  
Codex Marianus,  
passim



# Glagolitic Translit.

Needed: Cyr. Iota + dotless i

ll ii i x

- Iota ≠ duplicate form, not Greek
- Cyr. iota with dasia (asper) is freq. variant
- used by Leskien, Jagić, Diehls etc.
- compatibility with legacy data important
- i hardly in use instead of iota

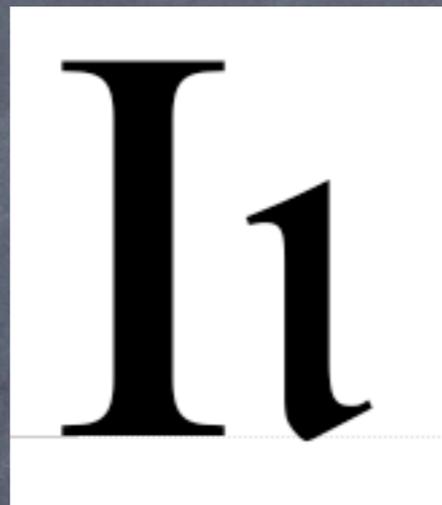


# Glagolitic Translit.

Design: Cyrillic Iota



OCS  
Black

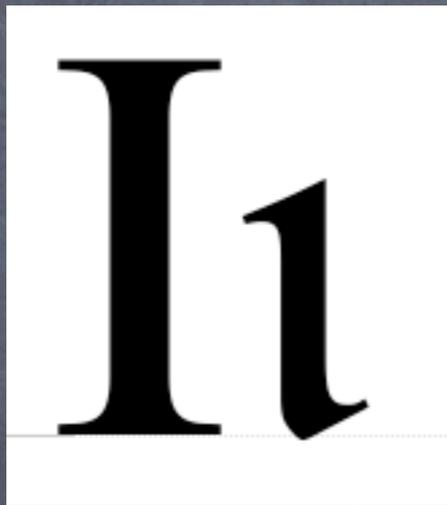


Greek  
Serifed

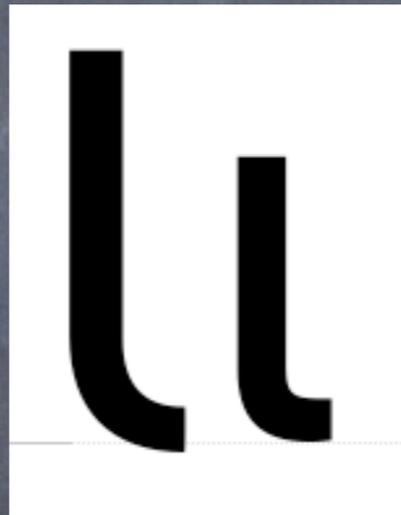


# Glagolitic Translit.

Design: Cyrillic Iota



Greek  
Serifed  
Roman



Cyrillic  
Sans  
Roman



Cyrillic  
Serifed  
Roman



# Glagolitic Translit.

Design: Cyrillic Iota

⋈ <sup>к</sup> в <sup>т</sup> по<sup>а</sup>не <sup>х</sup> е<sup>а</sup>тъ <sup>а</sup> не <sup>х</sup> е<sup>а</sup>тъи  
⋈ ап<sup>к</sup>лъ ⋈ ал<sup>е</sup> гла <sup>а</sup> ѿ ⋈  
Žalm 88. 6. <sup>т</sup> ѿповѣ<sup>д</sup>а <sup>а</sup> н<sup>б</sup>са <sup>а</sup> чю ⋈



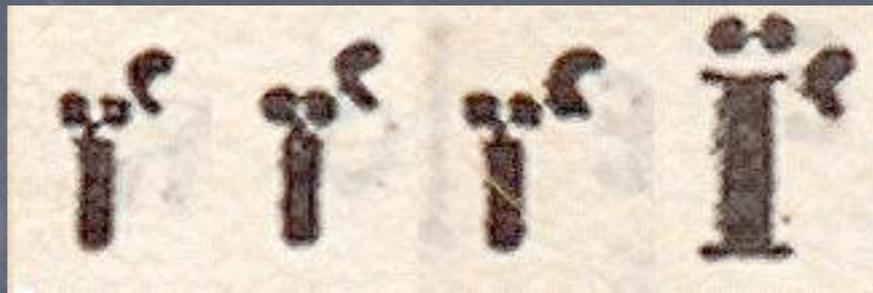
<sup>а</sup> не <sup>ѿ</sup> ѿ <sup>ѿ</sup> е<sup>в</sup> <sup>т</sup> ѿ <sup>ѿ</sup> ма <sup>ѿ</sup> гла <sup>ѿ</sup> р<sup>ѿ</sup>п<sup>ѿ</sup>ѣ ⋈

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



# Glagolitic Translit.

Transliteration with i



not widely used

Н. Каринский, Хрестоматия по древне-  
церковно-славянскому и русскому языкамъ.  
Часть первая. Древнейшие памятники. Изд. 2-е.  
С.-Петербургъ 1911, 3.



# Glagolitic Translit.

Transliteration with i



not widely used  
creates problems

for typesetting and interpretation

Н. Каринский, Хрестоматия по древне-  
церковно-славянскому и русскому языкамъ.  
Часть первая. Древнейшие памятники. Изд. 2-е.  
С.-Петербургъ 1911, 3.



# Glagolitic Translit.

## Transliteration of Nasals

€ → A x



# Glagolitic Translit.

## Transliteration of Nasals

Ѣ → Ā X

Ѧ → A X



# Glagolitic Translit.

## Transliteration of Nasals

Ѣ	→	А	х
Ѥ	→	А	х
Ѧ	→	Х	
Ѩ	→	Х	х



# Glagolitic Translit.

## Transliteration of Nasals

Ѣ → Ā x

Ѧ → A x

ѢѦ → X

ѦѢѦ → HX x

ѦѢ → 'YO' x



# Glagolitic Translit.

## Transliteration of Nasals

Ѣ → Ā x

Ѧ → A x

ѢѦ → X

ѦѢѦ → HX x

ѦѢ → ? ?



# Glagolitic Translit.

22

N. S. Trubetzkoy

I. Gruppe (Einer)			II. Gruppe (Zehner)			III. Gruppe (Hunderter)			IV. Gruppe (Tausender)		
1	⌘	<i>a</i>	10	⌘ ⌘	<i>i</i> <sub>1</sub>	100	⌘	<i>r</i>	1000	⌘	<i>č</i>
2	⌘	<i>b</i>	20	⌘	<i>i</i> <sub>2</sub>	200	⌘	<i>s</i>	2000	⌘	<i>š</i>
3	⌘	<i>v</i>	30	⌘	<i>h</i>	300	⌘	<i>t</i>	3000	⌘	<i>h</i>
4	⌘	<i>g</i>	40	⌘	<i>k</i>	400	⌘	<i>ü</i> <sub>2</sub>	4000	⌘	<i>z</i>
5	⌘	<i>d</i>	50	⌘	<i>l</i>	500	⌘	<i>f</i>	5000	⌘	<i>ä</i> ( <i>ʔ</i> )
6	⌘	<i>e</i>	60	⌘	<i>m</i>	600	⌘	<i>x</i> <sub>2</sub>	6000	⌘	<i>x</i> <sub>1</sub>
7	⌘	<i>ž</i>	70	⌘	<i>n</i>	700	⌘	<i>o</i> <sub>2</sub>	7000	⌘	<i>ö</i>
8	⌘	<i>z</i>	80	⌘	<i>o</i> <sub>1</sub>	800	⌘	<i>γ</i>	8000	⌘	<i>ü</i> <sub>1</sub>
9	⌘	<i>z</i>	90	⌘	<i>p</i>	900	⌘	<i>c</i>	9000	⌘	<i>N</i>



# Glagolitic Translit.

Ɀ	3
ⱿⱿ	ï
ⱿⱿ	za starije spomenike ġ, za mlade ĵ
Ɀ	ō
ⱿⱿ	ĉ <span style="color: green;">X</span>
ⱿⱿ	č
Ɀ	ju
ⱿⱿ	ž
Ɀ	6
Ɀ	ě, za vrijednost

?

Croatian  
Norm

ġ, ĵ X



# Croatian Glagoljica

Ѣ ѣ ѣ, ѣ ~~ѣ, ѣ~~? ѣ, ѣ  
Ь ь ь, ь; ь ~~ь, ь~~, ь

Different characters  
or stylistic variants?



# Bosančica



# Bosančica



ǰ	A	o	O
ǧ	B	п	P
п	V	р	R
ǎ	G	с	S
а, ђ, ѓ	D	ш	T
е	E	ѡ	U
ѡе, ж, ѡе	Ž	ф	F
ѣ	DZ	х	H
з	Z	ω	OT
н	I	ψ	Ć, ŠT, ŠĆ
ѣ, ѣ	Đ, Ć (đerv)	ч	C
ѡ	K	ѡ	Č
ѣ	L	ш	Š
ѣѣ	LJ	ѡ	poluglas
м	M	ѡ, ѡ	JAT
н	N	ю	JU
ѣн	NJ		



# Bosančica



ǰ	A	o	O
ǧ	B	п	P
п	V	р	R
л	G	с	S
а, ђ, ѓ	D	ш	T
е	E	ѱ	U
ѣ, ж, џ	Ž	ф	F
џ	DZ	х	H
з	Z	ω	OT
и	I	ψ	Ć, ŠT, ŠĆ
Ѡ, ѡ	Đ, Ć (đerv)	ϣ	C
к	K	ν	Č
л	L	ш	Š
љ	LJ	б	poluglas
м	M	ѿ, Ѡ	JAT
н	N	ю	JU
њ	NJ		



# Glagolica

“Pe Glagolitic” –  
Fact or Fiction?



# Glagolica

“Pe Glagolitic” –  
Facts

No text with this character



# Glagolica

## “Pe Glagolitic” – Facts

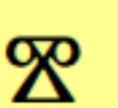
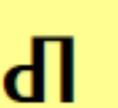
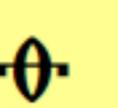
No text with this character  
Still Unicode has a slot for it  
– and a picture, too!

[Unicode 2C1A + 2C4A]



# Glagolica

“Pe Glagolitic” –  
Unicode Docs

	?		?			
A						
	2C0A	2C1A	2C2A	2C3A	2C4A	2C5A

2C18		GLAGOL
2C19		GLAGOL
2C1A		GLAGOL
2C1B		GLAGOL
2C1C		GLAGOL

Unicode Chart & Text





# Glagolica

Роуцѣ во свой горѣ | въздѣю присно .: |  
Силоу приати . и | моудрость оу те|бе .: |  
Ты во даиши . до|стойномъ си|лоу .: |  
Упостась же . вса||кою цѣлиши .: |  
Фарабша ма . зѣ|ловы и|збави .: |  
Херовьскоу ми . | мѣсль и оумъ | дождь .: |  
О чьстнаѣ . прѣ|стаѣ троице .: |  
Печаль мою на ра|дость прѣложи .: |  
Цѣломоудрьно . | да начьноу пь|сати .: |  
Чюдеса твоѣ . прѣ|дивнаѣ зѣло .: |  
Шестькрилатъ . си|лоу възпримъ .: |  
Шьствоую ны|нѣ . по слѣдоу оу|чителю .: |  
Имени юю . и дѣлоу || послѣдоуѣ .: |  
Ивѣ сътворю . ѣва|нгельско сло|во .: |  
Хвалоу възда|ѣ . трци въ вѣ|ствѣ .: |

The  
'Proof'

?



# 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
*	***	***	***	***	***	***	***	***	***	***	***	***	***
	◀	◊	/	◊		◊	◊	◊	◊	◊	◊	◊	◊

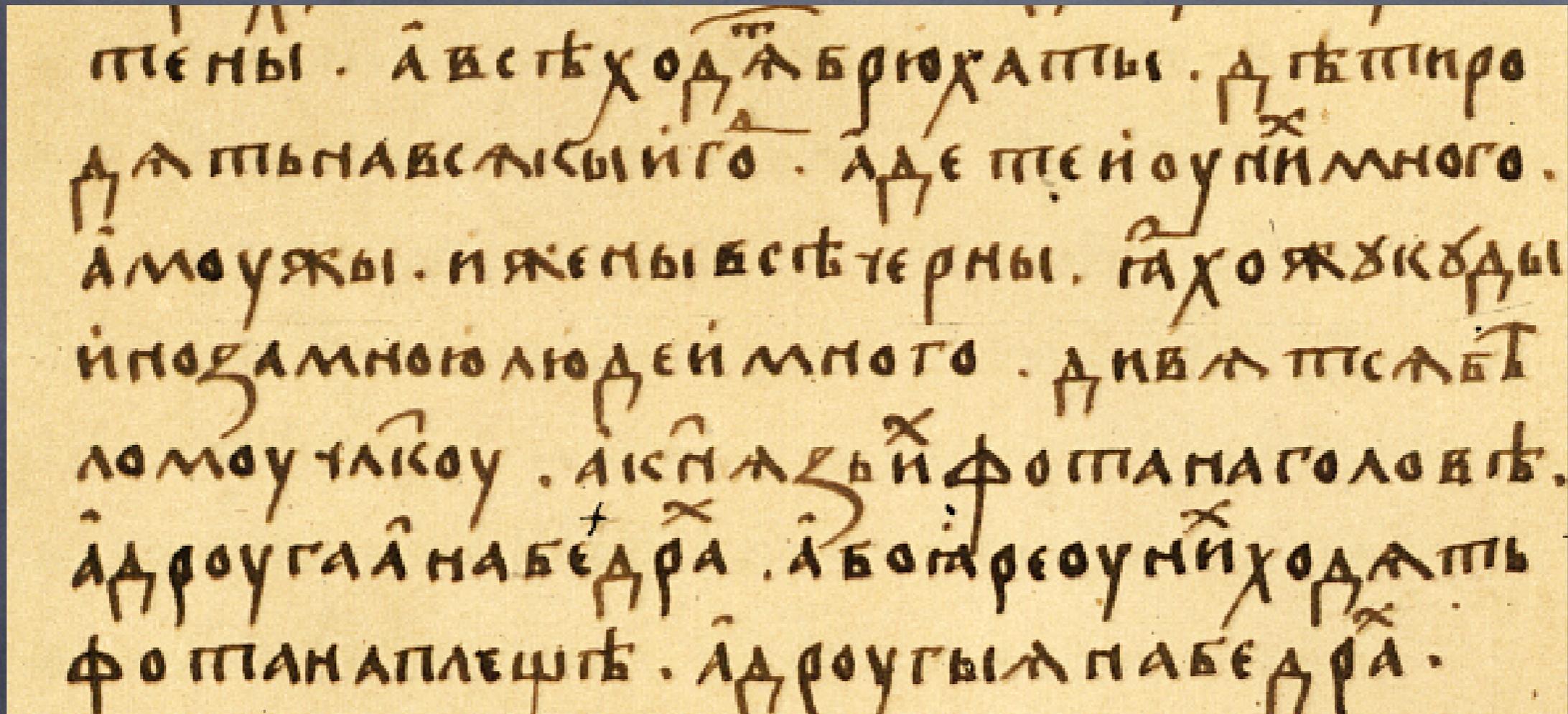
**Latin:** Many Superscripts (cmb)  
for Medievalists (German!):  
û (=ü) â (=ä) ô (=ö); v̇ etc.



(Comb. Diacritical Marks)



# Superscripts



Cyrillic: No Superscripts at all in UC



Афанасий Никитин, Хождение за три моря  
Троицкий список



# Ligatures

“Presentation forms”

- some for Latin
- many for Arabic

	FB0
0	<b>ff</b> FB00
1	<b>fi</b> FB01
2	<b>fl</b> FB02
3	<b>ffi</b> FB03
4	<b>ffl</b> FB04
5	<b>ft</b> FB05
6	<b>st</b> FB06



# Ligatures

	FB0
0	<b>ff</b> FB00
1	<b>fi</b> FB01
2	<b>fl</b> FB02
3	<b>ffi</b> FB03
4	<b>ffl</b> FB04
5	<b>ft</b> FB05
6	<b>st</b> FB06

“Presentation forms”

- some for Latin
- many for Arabic
- none for Cyrillic
- none for Glagolitic

ѢѢ  
ѦѦ  
ѦѦ  
ѦѦ  
ѦѦ  
ѦѦ  
ѦѦ  
ѦѦ



# Ligatures

## Ligatures

æ fb ct fy ee ff gi fh it fj fh it fj ky fl  
 gg gy oe sp ggy fr st ft ip py tw tt tw tty  
 AA MB CC MD ME FF CC HE FI UB NK FL LA  
 NT O MP E R W TT U V A TW UL TY UR  
 ffb ffi ffh ffj cky ffl ffr  
 fft ffy fi N js fs æ TE Æ

A modern Baskerville style type face set;  
 "Just Ligatures" from the "Mrs. Eaves" collection.  
 Zuzana Licko designer. Emigre.com

Unicode is not the  
 place to handle  
 ligatures – font  
 technology instead

XLVIII

LITTERAE IUNCTAE FREQUENTER OCCURRENTES

ab	av	ava	avv	ad	az	adz	ak	al	an	ap
at	at	aju	o	h2	bi	biv	bjud	bo	br	o
va	vv	vvr	vd	vdtl	v2	vzd	vz	vza	vzv	vzd
vzda	vzdov	vzli	vi	vla	vli	vjud	vo	vr	vrt	vt
vtr	o	ga	gv	gvl	gd	gda	gz	gl	gla	glv
giz	git	giu	go	gov	govl	goz	gr	gu	o	da
dv	di	o	eze	o	za	zv	zd	zda	ze	zi
zi	zju	o	za	zv	zvi2	zd	zda	z2	zi	zia
ziv	zmi	zo	zr	zra	o	iz	ize	il	ili	it
o	jl	o	ko	o	la	lv	lvi	lvt	lz	li
it	itv	itr	iju	jud	o	ma	m2	mzd	mi	mi
mitv	mo	o	nd	ndl	n2	no	o	od	o	pa
pv	pvd	pva	pvl	pl	plt	pml	po	pov	pod	poda
pot	pt	ptv	prt	pt	ptr	o	ra	rv	o	sa
al	st	o	ta	tv	tvo	tvd	tz	tl	tju	to
tr	trt	tt	o	ha	hv	h2	hl	ho	hov	hod
hot	hr	ht	hu	o	ta	ti	o	ju	ju2	juze
ju	ju	jut	ju2	o	o	o	o	o	o	o

Převládá: Ivan Šukman



# Balkan Philology

- OCS–Cyrillic for Romanian

## LITERA Л

Litera Л, obținută prin modificarea grafică a literei Ж<sup>78</sup>, în documentele muntenești ale secolului al XV-lea apare o singură dată, într-un act de întărire dat pentru mănăstirea Govora<sup>79</sup> de voievodul Radu cel Mare: **УТ ГЛАВ ПОИЕНЕ ДОЛЕ ПО ЛН<sup>80</sup> ДОЛННХ ЧОРЖЧЕВ** „din capul poienei, în jos, prin (pe în) valea Ciorîcei“ (1499 iul. 13, Arh. St. Buc., S. I., nr. 118).

Deși atestată într-un cuvânt 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<sup>81</sup>. În acest unic caz, grafia ne determină să afirmăm că Л notează nazalitatea vocalei care precedă pe Н<sup>82</sup>.

L. Djamo–Diaconiță: Limba documentelor slavo–române emise în Țara Românească în sec. XIV și XV. București 1971, 49



# Balkan Philology

- Greek for Albanian

## 3. Griechische Lettern.

Γιάτι ἴνε **κῆ** γέ μπῆ κίεῖ **κίεῖ** κιοφτε **σῆντε**ρούαρε ἔμερι ἴτ· ἄρτε μπρε-  
τερία γιότε· οὐ **πέφτε** οὐρῶρι ἴτ, σὶ κούντρο **πένε**τε ντῆ κίεῖ ἀστοῦ ἔ δὲ  
μπῆ δέ· ἔπνα νάθετ πούκῆν ἔ σόμμε **κῆ** νὰ δούχετε **πέρ** φῶστενε· ἔ δὲ

Transscription: *Yati inü tšü ye mbü kiel', kióftü süntüruarü 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 duçetü per fistünü. E de ndüléna*

# Balkan Philology

- Greek for Albanian

## 3. Griechische Lettern.

Γιάτι ἴνε **κῆ** γέ μπῆ κίεῖ **κίεῖ** κιοφτε **σῆντε**ροούαρε ἔμερι ἴτ· ὄρτε μπρε-  
τερία γιότε· οὐ **πέφτε** οὐρδερι ἴτ, σὶ κούντροε **πέντε** ντῆ κίεῖ ἀστοῦ ἐ δὲ  
μπῆ δέ· ἔπνα νάθετ πούκεν ἐ σόμμε **κῆ** νὰ δούχετε **πέρ** φούστενε· ἐ δὲ

Transscription: *Yati inü tšü ye mbü kiel', kióftü süntüruarü 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 duçetü per fistünü. E de ndüléna*

- Greek for Macedonian...

# Balkan Philology

- Greek for Albanian

## 3. Griechische Lettern.

Γιάτι ἴνε **κῆ** γέ μπῆ κίεῖ **κί** κιώφτε **σέντε**ρούαρε ἔμερι ἴτ· ὄρτε μπρε-  
τερία γιότε· οὐ **πέ**φτε οὐρῶρι ἴτ, σὶ κούντρο **πέν**ετῆ ντῆ κίεῖ ἀστοῦ ἔ δὲ  
μπῆ δέ· ἔπνα νάθετ πούκῆν ἔ σόμμε **κῆ** νὰ δούχετῆ πῆρ φῶστῆνε· ἔ δὲ

Transscription: *Yati inü tšü ye mbü kiel', kióftü süntüruarü 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 duçetü per fistünü. E de ndüléna*

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

# General Phonetics

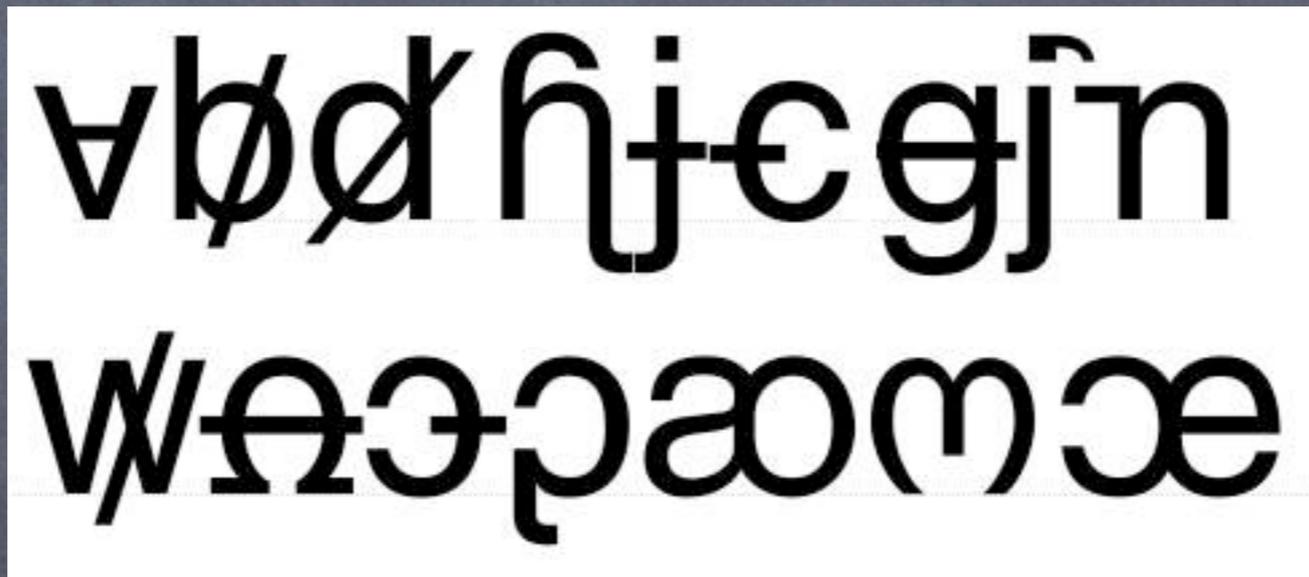
Not yet implemented...



Bavarian dialect atlas

# General Phonetics

Not yet implemented...



ʌ ɸ ø ħ j ɛ ɝ j ɾ  
ɰ ɛ ɜ ɝ ɹ ɶ ɹ ɔ ɯ ɶ

various phonetic symbols

# Medieval Philology

Not yet implemented...

ÉéŮůŮúaá̄Ōó



ʀʀʂʃ



gacute	AEacute	aeacute	Ostrokeac	ostrokeac	Adblgr.
á	Á	æ	Ø	ø	À
vertedb	einvrtedb	ldblgrave	idblgrave	linvrtedbl	invert
Ê	ê	Ï	ï	Î	î



# 4. Results



u

# Some Results...

kkk

lll

īō ḥa<sup>4</sup>

ov/ov/ov

lll

ÿ ÿ

Œ œ m̄

†

†

1D7C

ſ

ˆ ˆ

ŷ = dž

Handwritten symbol

≡ a ≡  
|||

7

€ A € A

æ



**Thank 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.**



**© Prof. Dr. Sebastian Kempgen 2021**

**ORCID: 0000-002-2534-9423**

**D-96045 Bamberg, University of Bamberg, Germany**

**[sebastian.kempgen@uni-bamberg.de](mailto:sebastian.kempgen@uni-bamberg.de)**

**<https://www.uni-bamberg.de/slaving/personal/prof-em-dr-sebastian-kempgen/>**

