Unicode for Slavic Medievalists

David J. Birnbaum
University of Pittsburgh
djbpitt+@pitt.edu

Euro Summer School:
Electronic Publishing for Cultural Heritage Studies

Sofia, 28 September 2002

Outline

- Lumpers and Splitters
- Early Cyrillic Writing and Unicode
- Early Glagolitic Writing and Unicode
- Where to Get Informatin and How to Participate

Lumpers and Splitters

- A few types of “e”: є ∈ ҽ ϵ ε э э ѥ
- Handwritten letter forms *always* differ
- Approaches to classification
  - *Splitter:* If they’re at all different, they’re different, and should be encoded differently
    - Example: Retain difference in rendering
  - *Lumper:* If they’re at all similar, they’re the same, and should be encoded identically
    - Example: Conflate differences for querying, collation

Early Cyrillic and Unicode

- “The historic form of the Cyrillic alphabet is treated as a font style variation of modern Cyrillic because the historic forms are relatively close to the modern appearance and because some of them are still in modern use … .”
- Early Cyrillic “𞀀” and modern Cyrillic “a” are both u+0430

Source Set Rule
Legacy Characters

Character	*Image*	*Decomposition*
u+0477	ѷ	u+0475 (ѵ) u+030f (◌̏)
u+047f	ѿ	u+0461 (ѡ) u+0442 (т) (superscript)
u+047d	ѽ	u+0461(ѡ) u+0483 (◌҃)

Complementary Distribution (Non-Slavic)

Language	*Phoneme*	*Spelling (lower case only)*
Greek	/s/	ς / __# (u+03c2) σ / elsewhere (u+03c3)
Hebrew	Final and nonfinal consonants	
Arabic	Initial, medial, final, and isolated consonants	

Complementary Distribution (Slavic)

Language	*Phoneme*	*Spelling (lower case only)*
Rusian (some)	/ja/ ~ /Ça/ (~ /Cä/)	ѧ / C __ ꙗ / elsewhere
Rusian (some)	/o/	ѻ / C __ ѡ / elsewhere
Pre-1918 Russian	/i/	і / __ V, й (exc. міръ) и / elsewhere

Visual Ambiguity

- Run-of-the-Mill Unicode Visual Ambiguity

  u+002d HYPEN-MINUS

  u+2010 HYPHEN

  u+2212 MINUS SIGN

- Super-Duper Early Cyrillic Visual Ambiguity
  - One character mapped to more than one glyph
  - One glyph associated with more than one character

  u+0486 COMBINING CYRILLIC PSILI PNEUMATA (’)

  u+0311 NON-SPACING INVERTED BREVE (̑)

Jotation

Character	*Image*	*Notes*
u+0465		
u+044f		
u+0469		
u+046d		
u+044e		Jotation plus u+043e () (!)
(none)		Jotation plus u+0463 ()
(none)		not present in Unicode Variant of u+0469 ()?

/i/

Character	*Image*	*Numerical*	*Notes*
u+0438	и	8	
u+0456	і	10	
u+0457	ї	10	
(none)	ı	10	Variant? Of what?
(none)	ɪ	10 (?)	Invented for transcriptions of Glagolitic

Cyrillic Palatal Glides

- U+0458 CYRILLIC SMALL LETTER JE (j)
  - Serbian, Macedonian
- U+0439 CYRILLIC SMALL LETTER SHORT I (й)
  - Russian, Ukrainian, Belarusian, Bulgarian

Jers

Character	*Image*	*Notes*
u+044a	Ъ	
u+044c	Ь	
(none)	ъ	“Neutral” jer Variant? Of what?

Jery

Part	*Images*
First	ъ, ь
Second	н, і, ї, ı, ι

- Non-ligated and ligated
- (Modern: u+044b ы)

Other Special Problems

- Upper and Lower Case
  - Modern and early Cyrillic are the same script
  - Modern Cyrillic distinguishes case
  - Early Cyrillic typically does not distinguish case
- Ligation is productive
- Superscription may require different glyphs
  - “Recumbant r” (ꙛ)

/u/

Character	*Image*	*Note*
u+0443	ѵ	Modern image y
u+0475	ѷ	Variously /ü/ or /u/
u+0479	ѹ	Horizontal digraph
u+043e u+0443	о ѵ	Sequence of two characters Second may alternatively be u+0475
(none)	ȣ	Vertical digraph (ligated)

/o/

Character	*Image*	*Notes*
u+043e	◌	Narrow
u+047b	◌	Broad
(none)	⊙⊙⊙	Ocular (also “polyocular”)

Cf. u+0461 ѡ (omega)

Cf. /e/

Greek

Greek		*Cyrillic*	
Character	*Image*	*Character*	*Image*
u+03c8	ψ	u+0471	ѱ
u+03be	ξ	u+046f	ѯ
u+03b8	θ	u+0473	ѳ
u+03b1	α	(none)	α
u+03b5	ε	(none)	ε

/e/

Character	*Image*	*Notes*
u+0435	е (modern)	CYRILLIC SMALL LETTER IE Most common modern *glyph*
u+0454	є (modern)	CYRILLIC SMALL LETTER UKRAINIAN IE Most common early *glyph*
u+044d	э (modern)	
u+0465	ѣ	
(none)	∈ ҿ ҽ ɛ ɜ ɜ	Variants? Of what?

Palatal Consonants

Character	*Image*	*Notes*
u+0459	љ (modern)	CYRILLIC SMALL LETTER LJE
u+045a	њ (modern)	CYRILLIC SMALL LETTER NJE
u+04a5	ҥ	CYRILLIC SMALL LIGATURE EN GHE (Altay, Mari, Yakut)
(none)	ʎ	Palatal /l/
u+04a5 (?)	ɲ	Palatal /n/
(none)	ʤ	Palatal /d/
(none)	(not available)	Palatal /m/

Old Church Slavonic and Rusian

Character	*Image*	*Sound*	
		Old Church Slavonic	*Rusian*
u+044f u+0469		/ę/ ~ /ję/	/ä/
u+0467		/ja/	

Front Nasal (Unicode)

Character	*Image*	*Notes*
u+0467		CYRILLIC SMALL LETTER LITTLE YUS
u+0469		CYRILLIC SMALL LETTER IOTIFIED LITTLE YUS

Front Nasal (Manuscripts) 1

Manuscript	*Nonjotated*	*Jotated*
Savvina Kniga	 , (rarely)	
Zograph Folia		
Suprasliensis Słuck Psalter		
Hilandar Folia		
Ostromir Gospel		
Preslav ceramic inscription		

Front Nasal (Manuscripts) 2

Manuscripts with a Single Front Nasal Letter

Manuscript	*Image*	*Notes*
Undol′skij Folia		
Cyrillic Macedonian Folium		twice (/C__)

 may represent etymological front or back nasal

Mixed Corpora: Geographic

Old Church Slavonic u+0467 (ѧ) may correspond to:

Character	*Image*	*Recension*
u+044f	ꙗ	Rusian
u+0454	є	Serbian
u+046b	ѫ	Middle Bulgarian

But: Not all Rusian ꙗ, Serbian є, and Middle Bulgarian ѫ correspond to one another.

Diachronic Paleography: ѧ and ꙗ

Image	*Character*	*Period*
ѧ	u+0467	Early
я	u+044f	Modern
ꙗ	(none)	Early

	East Slavic	*South Slavic*
Sound	я = ꙗ = ѧ	я = ꙗ (≠ ѧ)
Paleography	я < ѧ	

Glagolitic and Unicode

"The Unicode Standard regards Glagolitic as a *separate* script from Cyrillic, not as a font change from Cyrillic. This position is taken primarily because Glagolitic appears unrecognizably different from Cyrillic, and secondarily because Glagolitic has not grown to match the expansion of Cyrillic. The Glagolitic script is not currently supported by the Unicode Standard."

Proposed Glagolitic Inventory

- ⰰⰱⰲⰳⰴⰵⰶⰷⰸⰹⰺⰻⰼⰽⰾⰿⱀⱁⱂⱃⱄⱅ ⱆⱇⱈⱉⱊⱋⱌⱍⱎⱏⱐⱑⱒⱓⱔⱕⱖⱗⱘⱙⱚ
- P2, simple back jus, stapić, square ot.
- Generic short and long titlo
  - Deprecate u+0483 COMBINING CYRILLIC TITLO

Notes on Glagolitic 1

- Round and square glagolitic are different "typographic variants" (fonts)
- ⰸⰹ is two characters, not one
- Stapić
  - Historically a variant of ⰻ
  - Graphically distinct
  - Treated as distinct by those who work with square Glagolitic

Notes on Glagolitic 2

- Round and square ⰔerV
  - Historically the same letter
  - Different functions (Ⱄⰰ and j)
  - Separate characters?
- Additional diacritics or punctuation?
- Upper and lower case?

Commission

Special Commission to the Executive Council of the International Committee of Slavists for the Computer-Supported Processing of Slavic Manuscripts and Early Printed Books

Commission Projects

- Unicode Early Cyrillic Proposal
- Unicode Glagolitic Proposal
- Agreement (private standardization) of part of Unicode PUA for Slavic medievalist community

How to Participate

- ѻ҃бщежи́тїе Obštežitie Portal
  - http://www.ceu.hu/medstud/ralph/obsht.htm
- Mailing list for Early Slavic Written Sources
  - slav-mss-list@port.ac.uk
  - Subscription information available at Obštežitie
- Commission Web Site
  - http://clover.slavic.pitt.edu/~repertorium/commission/ (after 15 October 2002)
- Webmasters:
  - David J. Birnbaum: djbpitt+@pitt.edu
  - Ralph Cleminson: ralph.cleminson@port.ac.uk