


SCRIPTA

& -SCRIPTA

8–9/2010

e-SCRIPTA

Proposal for a unified encoding of Early Cyrillic glyphs in the Unicode Private Use Area

**A proposal prepared on behalf of the Commission
for Computer Processing of Slavic Manuscripts and Early
Printed Books to the International Committee of Slavists**

*Victor Baranov
David J. Birnbaum
Ralph Cleminson
Heinz Miklas
Achim Rabus*

Introduction

This paper proposes an encoding standard for certain early Cyrillic characters and glyphs that, for different reasons, are not yet, are unlikely to be, or will never be included in the Universal Character Set (UCS) of the Unicode Standard, but are nevertheless used by parts of the paleoslavistic community. In order to render these units in a standard-conformant way, there are three options¹:

1. Change fonts
2. Use the Unicode Private Use Area (PUA)
3. Make use of OpenType technology

In the current paper, the authors concentrate on option 2. We propose “a sort of microstandardization of a portion of the PUA”¹. The starting point for this project is the inventory of early Cyrillic originally proposed at the conference held in Belgrade on 15–17 October 2007, and published as part of its proceedings². While

¹ Birnbaum, Cleminson, Kempgen, Ribarov 2008: 177.

² Standardizacija 2009: 24–243.

as a proposal to expand the UCS this was widely considered unrealistic, a compromise reached during discussions among interested parties during the Fourteenth International Congress of Slavists in Ohrid, 10–16 September 2008, proposed placing the additional units at agreed codepoints within the PUA, on the model of what had already been done for such elements used in medieval Latin-script texts by the Medieval Unicode Font Initiative (MUFI). The advantage of this is that it complies with the standard while at the same time giving medievalists the possibility of rendering texts exactly as they wish and maintaining compatibility with one another’s work.

Recent history

In recent years there have been lively discussions regarding the appropriate way to encode early Cyrillic writing. Some scholars in the field, including some of the authors of the current paper, follow a “maximalist” approach: they wish to reproduce the graphemic peculiarities of early Cyrillic texts as faithfully as possible. In doing so, they need to use graphemic entities that are not encoded in the Unicode standard. However, as was stated³, because of the Unicode Consortium’s encoding policy many of the units proposed in the Belgrade Standard cited above are not candidates for inclusion in the Unicode standard. In view of this situation, the colleagues adhering to the “maximalist” approach proposed encoding all the needed units – for systematic reasons including also the ones already present in the UCS—separately, although they were aware that creating such a parallel encoding strategy alongside the Unicode standard would be fraught with problems.

Other scholars, including some of the authors of the current paper, prefer a “minimalist” approach, i.e., they adhere strictly to the Unicode standard, use markup to represent details that cannot be encoded at the character level, and refrain from reproducing graphemic peculiarities in great detail. However, the impossibility of precisely rendering all graphemic peculiarities of a text is a serious disadvantage of the “minimalist” approach. The fundamental problem is that “there is no universal set of fundamental units of text. Instead, the division of text into text elements necessarily varies by language and text process⁴.” In the present case, the minimalist approach favors the text processes of searching and sorting, so that letter forms that represent the same linguistic unit will be encoded similarly.⁵ The maximalist approach favors the text process of rendering, ensuring that letter forms that look different in original sources will be encoded differently, and therefore can easily be rendered differently.

³ Birnbaum, Cleminson, Kempgen, Ribarov 2008

⁴ US 5.2.0, Chapter 2.

⁵ For an earlier attempt to reconcile the inherent conflicts among text processes within the Unicode framework, see (Birnbaum, Cournane, Flynn 1999).

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

The authors of the current paper seek to reconcile the “maximalist” and “minimalist” positions. They recommend a general convention, or micro-standard, for those glyphs that are needed for specific texts, but will not find their way into the UCS. The proposed micro-standard is meant to expand, not to replace the Unicode standard. An easy, practical and convenient way to achieve the aim of a micro-standard of early Cyrillic writing is to coordinate the use of a portion of the Unicode Private Use Area (PUA) for the purpose of encoding glyphic variants of early Cyrillic characters on a character level.

Why do we need early Cyrillic PUA microstandardization?

There are three recommended options for Slavic medievalists who need to deal with letter forms that are not included in the inventory of Unicode⁶:

1. Change fonts. Use a set of fonts that are all Unicode-encoded but contain differently shaped glyphs. This option can be useful in some cases in order to represent the writing of each “time and place in a culturally acceptable manner”⁵. The font solution, however, uses a presentational technology (font) for informational purposes, and therefore puts all users of the document at risk of losing information should they not have access to the original fonts, since other sets of fonts may have differently shaped glyphs. Furthermore, choosing the font solution means that one can neither add any characters missing from the Unicode inventory nor achieve independence from fonts and, thus, extensive compatibility. One will most likely run into problems if one tries to exchange one’s data with other users who may not have access to the same project-specific fonts.

2. Use the Unicode Private Use Area. If the paleoslavistic community is able to reach an agreement about where in the PUA to place the required units that are not present in the Unicode standard, thus establishing a sort of microstandardization, scholars can adhere to standards, ensuring sustainability and compatibility of their texts and, at the same time, render texts as exactly as they wish.

3. Use OpenType technology. OpenType is a technology that allows multiple glyphs to be assigned to one code point. It also provides graceful support for ligatures (explicitly excluded from Unicode). Some promising possibilities of OpenType have been demonstrated by Kempgen and Rabus⁷. However, for the time being there are few applications supporting OpenType features, which means that OpenType technology cannot currently be considered a viable alternative⁸.

⁶ Birnbaum, Cleminson, Kempgen, Ribarov 2008: 177

⁷ Kempgen 2008; Rabus 2010.

⁸ Adobe InDesign, a mid-level end publishing and layout application used primarily by professional publishers, provides excellent support for OpenType, but most Slavists work entirely within word processors and other applications that are designed to support

Thus, the most promising of the options seems to be solution 2, i.e., to use the Unicode Private Use Area (PUA). At the moment, the use of the PUA provides the only convenient and technically feasible compromise between the “maximalist” and “minimalist” positions.

Coordination with MUF1

The present paper is not the first proposal by medievalists to create in the PUA a sort of micro-standardization of units not included in the Basic Multilingual Plane of the Unicode Character Set. For medieval characters and glyphs of the Latin script, the Medieval Unicode Font Initiative (MUF1) elaborated a PUA-standardization known as the MUF1 character recommendation⁹.

The authors of the present paper seek to enable scholars to use Latin and Cyrillic PUA characters at the same time. Hence, the proposed Cyrillic characters and glyphs have been assigned to code points of the PUA that are not occupied by MUF1 characters. Medievalists working on Latin and Cyrillic texts in parallel should thus not run into encoding problems, provided they adhere to both the MUF1 character recommendation and the new Cyrillic additions. For the latter we will use the abbreviation CYFI.

According to Stötzner¹⁰, one relatively large PUA block not occupied by MUF1 characters is F330–F3FF. Additionally, the block EF60–EF8F is completely unused. We propose that these blocks be used for the Early Cyrillic characters and glyphs described in the present paper.

Principles for inclusion of characters/glyphs

It goes without saying that PUA characters and glyphs should not replace proper Unicode characters; rather they should be used exclusively as additions to those characters. As our colleagues from MUF1 rightly state: “Characters in the Private Use Area (PUA) should be used with great caution. [...] For documents with a long life expectancy, it is strongly recommended that PUA characters should be encoded with mark-up or entities, and that PUA characters should be used for the final display only, whether on screen or in print. For documents with a short life expectancy, characters may be used with less caution, as long as future problems of storage and interchangeability are considered.”¹¹

composition, rather than layout. The latest version of Microsoft Word for Windows (2010) provides limited support for OpenType features such as ligatures or style sets. Of current technologies, OpenType is The Correct Solution to the problem we are addressing, but the lack of support for it in the mainstream applications used by most paleoslavists means that an alternative approach is required, at least in the interim.

⁹ See <http://www.hit.uib.no/mufi/>.

¹⁰ Stötzner 2009

¹¹ MUF1 character recommendation 2009

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

Due to this issue, the authors of the present paper resolved to include only as many units as absolutely required. This is reflected in the principles for inclusion of characters and glyphs.

The following symbols have been included:

1. Functional characters such as CYRILLIC LETTER BROAD ES (capital, small and combining characters)
2. Mirrored characters such as CYRILLIC LETTER REVERSED A
3. Composite characters that can be rendered as sequence of a Unicode character and a diacritic Unicode sign, such as CYRILLIC LETTER PALATAL GHE (capital, small and combining characters). The reason for including these characters is to guarantee correct typographic placement
4. Composite characters that can be rendered as a sequence of two Unicode characters, such as CYRILLIC LETTER YERU WITH BACK YER AND IOTA¹². These digraphic units are mainly needed for transliteration (Cyrillic-Glagolitic) and sorting purposes
5. Superscript characters not (yet) present in the Base Multilingual Plane (BMP, code points between U+0000 and U+FFFF) of the UCS. There are some superscript characters currently considered for inclusion to the BMP¹³. Due to the fact that it is not yet known if and when all super-script characters will be included in the BMP, they have temporarily been included in the proposal at hand, although we recommend that their use be de-precated should they be included in a subsequent revision of the Unicode standard. Examples include COMBINING CYRILLIC LETTER HARD SIGN
6. Modifier characters such as COMBINING CYRILLIC INVERTED TITLO.
7. Punctuation marks such as REVERSED THREE DOT PUNCTUATION

The following symbols have not been included in the present proposal:

8. Glyph variants, the reproduction of which is of limited interest to paleo-slavists
9. Ligatures that represent combinations of sounds
10. Combining (superscript) characters consisting of a sequence of two characters in vertical order

¹² The name YERU (for YERY) is plainly erroneous, but the ISO standardization process requires that, in the interest of easing migration, established character names, even if plainly erroneous, not be changed. However odd the word “yeru” may look to a Slavist, what is important for the migration of legacy documents from earlier to later standards is the ability to map unambiguously from what used to be called “yeru,” a process facilitated by perpetuating the otherwise unfortunate name in the newer standards.

¹³ Everson, Baranov, Miklas, Rabus 2010

Correction signs

Correction signs are presently lacking in Unicode. They can be used for transcriptions of illegible characters or symbols that for various reasons have not been preserved in the originals (see Modifier Characters). So far these symbols have been encoded variously in critical and diplomatic editions, cf., e.g., the New Testament editorial symbols (2E00–2E0D) and the Ancient Greek textual symbols (2E0E–2E16).

Explanations

The units needed for the electronic rendering of Old Slavonic texts can be divided into three groups:

(1) main units and their combinations, needed for a simplified rendering in plain text, that is, without markup,

(2) functional variants and their combinations, needed for the exact graphemic reproduction of documents,

(3) paleographic variants (glyphs) and their combinations, needed for a formally exact reproduction of manuscripts and early printed material.

Clearly this division is to be taken *cum grano salis*, as it depends on the exact spatiotemporal, sometimes even individual, use of the given unit. Therefore, each choice and attachment of a unit to a certain group would have to be explained separately (for the gross differentiation of functional units¹⁴. Presently, Unicode 5.2 contains all units of the group (1) and some units of the group (2).

Since the number of free code points in MUFI is limited, here we propose to include only symbols (characters, superscript characters) with a specific function and coordinate their location in PUA with MUFI. This set we will call PUA1. In the future a second set PUA2 will be proposed for a number of glyphs (ligatures, paleographic variants) that may not be coordinated with MUFI and would therefore be intended for Slavistic projects that will not also require support for MUFI encoding.

Thus, the present choice foresees:

1. to include in PUA1 character symbols of group (2), including:

a. variants of symbols in the Unicode standard:

i). mirror variants, e.g. CYRILLIC LETTER REVERSED A,

ii). palatal and soft variants, e.g. CYRILLIC CAPITAL LETTER SOFT EN,⁵

¹⁴ Baranov, Romanenko 2009

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

iii). superscript letters, e.g. COMBINING CYRILLIC LETTER BROAD IE, in this section also:

- superscript variants of modern character forms that are needed for the transliteration of old texts with modern means, e.g. COMBINING CYRILLIC LETTER KJE,
- superscript characters with diacritic, for which the standard foresees combinations, e.g. COMBINING CYRILLIC LETTER YI;

b. to include character symbols needed for the exact transliteration of Glagolitic texts, comprising also superscript letters such as COMBINING CYRILLIC LETTER IOTA;

2. to include into PUA1 symbols of the group (3) that differ notably from the existing variants in Unicode, e.g. CYRILLIC LETTER REVERSED REARRANGED UK;

3. to include in PUA1 modifying symbols (titlo, paerok, diacritics such as accents or spiritus), e.g. COMBINING CYRILLIC INVERTED TITLO, punctuation signs, e.g., SLAVONIC KOPYE;

4. to include in PUA1 correction signs;

5. to refrain from including in PUA1

- a. symbols of the group (3) with the exception of those mentioned above, which are to be included into PUA2;
- b. combinations, with the exception of those mentioned above, i.e., combinations of characters with titlo or diacritics, ligatures and characters denoting numbers. They will be included in PUA2;
- c. certain theoretically possible symbols that are of limited interest to paleoslavists, e.g. CYRILLIC LETTER SOFT ER;

6. to refrain from leaving spare code numbers after every group to be used for new symbols of the given group.

7. to refrain from leaving spare code numbers after every group to be used for new symbols of the given group.

It is hoped that the proposed PUA encoding for Early Cyrillic Symbols will establish itself as a sort of micro-standardization. Designers of scholarly fonts are encouraged to include characters and glyphs according to our proposal (see code points in appendix).

REFERENCES

- Baranov, Romanenko 2009*: Баранов, В. А., Романенко В.А., Опыт разработки, создания и использования кирилловского алфавита для полнотекстовых баз данных и интернет-изданий древнерусских рукописей XI–XIV веков. // Јовановић Г., Грковић-Мејџор Ј., Костић З., Савић В. (уред.) *Стандардизација старословенског ћириличног писма и његова регистрација у Уникоду*. Београд, 2009 (= САНУ, Научни скупови, Књ. СХХV; Одељење језика и књижевности, књ. 20), 49–62.
- Birnbaum, Cournane, Flynn 1999*: Birnbaum D. J., Cournane M., Flynn P., Using the TEI Writing System Declaration (WSD). // *Computers and the Humanities*, 1999, 33/1–2 (April), 49–57.
- Birnbaum, Cleminson, Kempgen, Ribarov 2008*: Birnbaum D. J., Cleminson R., Kempgen S., Ribarov K., Character Set Standardization for Early Cyrillic Writing after Unicode 5.1 (A White Paper prepared on behalf of the Commission for Computer Processing of Slavic Manuscripts and Early Printed Books to the International Committee of Slavists). // *Scripta & e-Scripta* 6, 2008, 161–193.
- BP: Belgrade Proposal – Стандард старословенског ћириличног писма / коначни предлог*. Миклас Х., Баранов В. А., Костић З., Савић В. (аутори). – Света Гора Атонска, Манастир Хиландар (= Monastery Hilandar), 2008 Београд, ИЦА, 24 стр. – ISBN 978-86-84747-30-5.
- Everson, Baranov, Miklas, Rabus 2010*: Everson M., Baranov V., Miklas H., Rabus A., *Proposal to encode nine Cyrillic characters for Slavonic*: <http://std.dkuug.dk/jtc1/sc2/wg2/docs/n3748.pdf>
- IRSoFCS: Kostić Z., Internal registration of the Standard of OCS (Belgrade Model) in Unicode PrivateUseArea (PUA)*. Belgrade, 2008, 93 pp.
- Kempgen 2008*: Kempgen S., Unicode 5.1, Old Church Slavonic, Remaining Problems – and Solutions, including OpenType Features. // Miklas H., Miltenova A. (eds.), *Slovo: Towards a Digital Library of South Slavic Manuscripts. Proceedings of the International Conference, 21–26 February 2008, Sofia, Bulgaria*. Sofia 2008, 200–219. Online version: http://kodeks.uni-bamberg.de/slavling/downloads/SK_Unicode_5.1_OCS_OTF.pdf
- MUFI character recommendation 2009*: MUFI character recommendation. Characters in the official Unicode Standard and in the Private Use Area for Medieval texts written in the Latin alphabet. Version 3.0 of 5 July 2009. <http://www.mufl.info/specs/MUFI-Alphabetic-3-0.pdf>
- Rabus 2010*: Rabus A., Unicode and OpenType – a practical approach to producing Church Slavonic scientific editions. // Баранов В. А. (отв. ред.), *Письменное наследие и современные информационные технологии: сборник статей лекторов междунар. науч. школы для молодежи (Ижевск, 12–15 октября 2009 г.)* Ижевск, 2010 (in press).

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

Standardizacija 2009: Јовановић Г., Грковић-Мејџор Ј., Костић З., Савић В., (уред.), *Стандардизација старословенског ћириличног писма и његова регистрација у Уникоду*. Београд, 2009 (= САНУ, Научни скупови, Књ. СХХV; Одељење језика и књижевности, књ. 20).

Stötzner 2009: Stötzner A., A Private Use Area Survey. A comparative study of the Privat [!] Use Area of the Unicode Standard as used by fonts and projects in the fields of Indo-European linguistics. Part One: Detailed Version.
http://www.signographie.de/cms/upload/PUA%20_%20SIAS/SIAS_PUA_survey_in_detail.pdf

US 5.2: The Unicode Standard, Version 5.2. Ranges: 0400–04FF, 2DE0–2DFF, A640–A69F (URL: <http://unicode.org>).

Appendix

Structure of tables

- 1 – Character name
- 2 – Indication: Upper-Case Character (UC), Lower-Case Character (LC), Superscript Character (SC), Upper-Case Ligature Characters (ULC), Lower-Case Ligature Characters (LLC), Superscript Ligature Characters (SLC)
- 3 – New Character code point in PUA
- 4 – Glyph
- 5 – Code of the Belgrade Proposal
- 6 – Code of the Internal registration of the Standard of OCS

Number of units in the Private Use Area 1

Upper-Case Characters (UC) – 38.

Lower-Case Characters (LC) – 35.

Superscript Characters (SL) – 88.

Modifier Characters (MC) – 9.

Punctuation Characters (PC) – 7.

In total – 177 characters.

Tables

<i>Characters</i>					
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
CYRILLIC CAPITAL LETTER REVERSED A	UC	F330	Ɑ	-	-
CYRILLIC SMALL LETTER REVERSED A	LC	F331	Ɱ	-	-
COMBINING CYRILLIC LETTER REVERSED A	SC	F332	Ɑ̂	-	-
CYRILLIC CAPITAL LETTER PALATAL GHE	UC	F333	Ɱ̂	5	1009
CYRILLIC SMALL LETTER PALATAL GHE	LC	F334	Ɱ̃	5	1019
COMBINING CYRILLIC LETTER PALATAL GHE	SC	F335	Ɱ̂̃	5	1011
COMBINING CYRILLIC LETTER GIE	SC	F336	Ɱ̃̂	-	-
COMBINING CYRILLIC LETTER SOFT DE	SC	F337	Ɱ̃̆	7	1108
COMBINING CYRILLIC LETTER IO	SC	F338	Ɱ̃̈	-	-
COMBINING CYRILLIC LETTER UKRAINIAN IE	SC	F339	Ɱ̃̉	9	1198
CYRILLIC CAPITAL LETTER EPSILON	UC	F33A	Ɱ̃̊	-	1129
CYRILLIC SMALL LETTER EPSILON	LC	F33B	Ɱ̃̋	-	1162
COMBINING CYRILLIC LETTER EPSILON	SC	F33C	Ɱ̃̌	-	1137
COMBINING CYRILLIC LETTER E	SC	F33D	Ɱ̃̍	10	1237
COMBINING CYRILLIC LETTER DZELO	SC	F33E	Ɱ̃̎	12	1323
COMBINING CYRILLIC LETTER DZE	SC	F33F	Ɱ̃̏	13	1360
COMBINING CYRILLIC LETTER REVERSED DZE	SC	F340	Ɱ̃̐	14	1394
COMBINING CYRILLIC LETTER ZEMLYA	SC	F341	Ɱ̃̑	16	1471
COMBINING CYRILLIC LETTER I	SC	F342	Ɱ̃̒	17	1514
COMBINING CYRILLIC LETTER SHORT I	SC	F343	Ɱ̃̓	-	-
COMBINING CYRILLIC LETTER BYELORUSSIAN-UKRAINIAN I WITH DOT	SC	F344	Ɱ̃̔	-	-
CYRILLIC SMALL LETTER DECIMAL I	LC	F345	Ɱ̃̕	19	1614
COMBINING CYRILLIC LETTER DECIMAL I	SC	F346	Ɱ̖̃	19	1598

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

COMBINING CYRILLIC LETTER YI	SC	F347	Ѣ̃	-	-
COMBINING CYRILLIC LETTER IOTA	SC	F348	Ѣ̇	22	1722
COMBINING CYRILLIC LETTER YE	SC	F349	Ѣ̆	-	-
CYRILLIC CAPITAL LETTER DJERV WITHOUT STROKE	UC	F34A	Ѣ̂	23	1754
CYRILLIC SMALL LETTER DJERV WITHOUT STROKE	LC	F34B	ѣ̂	23	1773
COMBINING CYRILLIC LETTER DJERV WITHOUT STROKE	SC	F34C	Ѣ̃	23	1761
COMBINING CYRILLIC LETTER MODERN DJERV	SC	F34D	Ѣ̆	25	1835
CYRILLIC CAPITAL LETTER PALATAL KA	UC	F34E	Ѣ̂	27	1915
CYRILLIC SMALL LETTER PALATAL KA	LC	F34F	ѣ̂	27	1930
COMBINING CYRILLIC LETTER PALATAL KA	SC	F350	Ѣ̃	27	1918
COMBINING CYRILLIC LETTER KJE	SC	F351	Ѣ̆	-	-
COMBINING CYRILLIC LETTER SOFT EL	SC	F352	Ѣ̇	29	1995
COMBINING CYRILLIC LETTER LJE	SC	F353	Ѣ̆	-	-
COMBINING CYRILLIC LETTER SOFT EM	SC	F354	Ѣ̇	31	2087
CYRILLIC CAPITAL LETTER SOFT EN	UC	F355	Ѣ̂	33	2150
CYRILLIC SMALL LETTER SOFT EN	LC	F356	ѣ̂	33	2165
COMBINING CYRILLIC LETTER SOFT EN	SC	F357	Ѣ̃	33	2153
COMBINING CYRILLIC LETTER NJE	SC	F358	Ѣ̆	-	-
CYRILLIC CAPITAL LETTER BROAD O	UC	F359	Ѣ̂	35	2215
CYRILLIC SMALL LETTER BROAD O	LC	F35A	ѣ̂	35	2236
COMBINING CYRILLIC LETTER BROAD O	SC	F35B	Ѣ̃	35	2224
COMBINING CYRILLIC LETTER MONOCULAR O	SC	F35C	Ѣ̇	36	2267
COMBINING CYRILLIC LETTER BINOCULAR O	SC	F35D	Ѣ̆	37	2305
COMBINING CYRILLIC LETTER DOUBLE MONOCULAR O	SC	F35E	Ѣ̇	38	2340
COMBINING CYRILLIC LETTER KOPPA	SC	F35F	Ѣ̆	-	-
CYRILLIC CAPITAL LETTER BROAD ES	UC	F360	Ѣ̂	43	2530
CYRILLIC SMALL LETTER BROAD ES	LC	F361	ѣ̂	43	2547

COMBINING CYRILLIC LETTER BROAD ES	SC	F362	ѐ	43	2535
COMBINING CYRILLIC LETTER TSHE	SC	F363	ћ	-	-
COMBINING CYRILLIC LETTER U	SC	F364	љ	48	2811
COMBINING CYRILLIC LETTER SHORT U	SC	F365	њ	-	-
COMBINING CYRILLIC LETTER STRAIGHT U	SC	F366	ћ	-	-
COMBINING CYRILLIC LETTER STRAIGHT U WITH STROKE	SC	F367	ќ	-	-
CYRILLIC CAPITAL LETTER REVERSED U	UC	F368	ѝ	-	-
CYRILLIC SMALL LETTER REVERSED U	LC	F369	ў	-	-
COMBINING CYRILLIC LETTER REVERSED U	SC	F36A	џ	-	-
COMBINING CYRILLIC LETTER UK	SC	F36B	Ѡ	45	2650
CYRILLIC CAPITAL LETTER REVERSED REARRANGED UK	ULC	F36C	ѡ	-	-
CYRILLIC SMALL LETTER REVERSED REARRANGED UK	LLC	F36D	Ѣ	-	-
COMBINING CYRILLIC LETTER REVERSED REARRANGED UK	SLC	F36E	ѣ	-	-
CYRILLIC CAPITAL LETTER REARRANGED UK	ULC	F36F	Ѥ	-	-
CYRILLIC SMALL LETTER REARRANGED UK	LLC	F370	ѥ	-	-
COMBINING CYRILLIC LETTER REARRANGED UK	SLC	F371	Ѧ	-	-
CYRILLIC CAPITAL LETTER LIGATURE UK	ULC	F372	ѧ	47	2749
CYRILLIC SMALL LETTER LIGATURE UK	LLC	F373	Ѩ	47	2778
CYRILLIC CAPITAL LETTER INVERTED LIGATURE UK	ULC	F374	ѩ	-	-
CYRILLIC SMALL LETTER INVERTED LIGATURE UK	LLC	F375	Ѫ	-	2724
CYRILLIC CAPITAL LETTER INSCRIBED UK	ULC	F376	ѫ	74	5062
CYRILLIC SMALL LETTER INSCRIBED UK	LLC	F377	Ѭ	52	5339
COMBINING CYRILLIC LETTER EF	SC	F378	ѭ	49	2844
CYRILLIC CAPITAL LETTER PALATAL HA	UC	F379	Ѯ	51	2920
CYRILLIC SMALL LETTER PALATAL HA	LC	F37A	ѯ	51	2935
COMBINING CYRILLIC LETTER PALATAL HA	SC	F37B	Ѱ	51	2923
COMBINING CYRILLIC LETTER OMEGA	SC	F37C	ѱ	52	2961

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

COMBINING CYRILLIC LETTER BROAD OMEGA	SC	F37D	Ѡ	-	3005
CYRILLIC CAPITAL LETTER CLOSED OMEGA	UC	F37E	Ѣ	54	3043
CYRILLIC SMALL LETTER CLOSED OMEGA	LC	F37F	ѣ	54	3061
COMBINING CYRILLIC LETTER CLOSED OMEGA	SC	F380	Ѣ̂	54	3049
COMBINING CYRILLIC LETTER ROUND OMEGA	SC	F381	Ѥ	55	3082
COMBINING CYRILLIC LETTER REVERSED TSE	SC	F382	Ѧ	59	3232
COMBINING CYRILLIC LETTER DZHE	SC	F383	ѧ	62	3343
CYRILLIC CAPITAL LETTER SHTA	ULC	F384	Ѩ	64	3400
CYRILLIC SMALL LETTER SHTA	ULC	F385	ѩ	64	3420
COMBINING CYRILLIC LETTER SHTA	SLC	F386	Ѩ̂	64	3408
CYRILLIC CAPITAL LETTER INVERTED SHTA	ULC	F387	Ѫ	64	3401
CYRILLIC SMALL LETTER INVERTED SHTA	ULC	F388	ѫ	64	3421
COMBINING CYRILLIC LETTER HARD SIGN	SC	F389	Ѭ	66	3486
COMBINING CYRILLIC LETTER YERU WITH BACK YER	SC	F38A	Ѯ	67	3532
CYRILLIC CAPITAL LETTER LIGATED YERU WITH BACK YER	UC	F38B	ѯ	-	3527
CYRILLIC SMALL LETTER LIGATED YERU WITH BACK YER	LC	F38C	Ѱ	-	3553
COMBINING CYRILLIC LETTER LIGATED YERU WITH BACK YER	SC	F38D	ѯ̂	-	-
CYRILLIC CAPITAL LETTER YERU WITH BACK YER AND IOTA	UC	F38E	Ѳ	69	3608
CYRILLIC SMALL LETTER YERU WITH BACK YER AND IOTA	LC	F38F	ѳ	69	3623
COMBINING CYRILLIC LETTER YERU WITH BACK YER AND IOTA	SC	F390	Ѳ̂	69	3611
CYRILLIC CAPITAL LETTER YERU WITH BACK YER AND I	UC	F391	Ѵ	70	3638
CYRILLIC SMALL LETTER YERU WITH BACK YER AND I	LC	F392	ѵ	70	3653
COMBINING CYRILLIC LETTER YERU WITH BACK YER AND I	SC	F393	Ѵ̂	70	3641
COMBINING CYRILLIC LETTER YERU	SC	F394	Ѷ	72	3708
CYRILLIC CAPITAL LETTER LIGATED YERU	UC	F395	ѷ	-	3703
CYRILLIC SMALL LETTER LIGATED YERU	LC	F396	Ѹ	-	3721

COMBINING CYRILLIC LETTER LIGATED YERU	SC	F397	Ѣ	-	-
CYRILLIC CAPITAL LETTER YERU WITH IOTA	UC	F398	Ѣ	74	3769
CYRILLIC SMALL LETTER YERU WITH IOTA	LC	F399	ѣ	74	3784
COMBINING CYRILLIC LETTER YERU WITH IOTA	SC	F39A	Ѣ̇	74	3772
CYRILLIC CAPITAL LETTER YERU WITH I	UC	F39B	ѢИ	75	3799
CYRILLIC SMALL LETTER YERU WITH I	LC	F39C	Ѣи	75	3814
COMBINING CYRILLIC LETTER YERU WITH I	SC	F39D	ѢИ̇	75	3802
COMBINING CYRILLIC LETTER SOFT SIGN	SC	F39E	ѣ̆	71	3671
COMBINING CYRILLIC LETTER NEUTRAL YER	SC	F39F	ѣ̇	76	3834
COMBINING CYRILLIC LETTER IOTIFIED YAT	SC	F3A0	Ѥ̇	78	3930
COMBINING CYRILLIC LETTER REVERSED YU	SC	F3A1	Ѥ̇	80	4004
CYRILLIC CAPITAL LETTER YUK	UC	F3A2	Ѥ	81	4035
CYRILLIC SMALL LETTER YUK	LC	F3A3	ѥ	81	4050
COMBINING CYRILLIC LETTER YUK	SC	F3A4	Ѥ̇	81	4038
CYRILLIC CAPITAL LETTER SEMIIOTIFIED A	UC	F3A5	Ѧ	83	4103
CYRILLIC SMALL LETTER SEMIIOTIFIED A	LC	F3A6	ѧ	83	4118
COMBINING CYRILLIC LETTER SEMIIOTIFIED A	SC	F3A7	Ѧ̇	83	4106
COMBINING CYRILLIC LETTER IOTIFIED E	SC	F3A8	Ѩ̇	84	4141
CYRILLIC CAPITAL LETTER REVERSED IOTIFIED E	UC	F3A9	Ѩ	-	-
CYRILLIC SMALL LETTER REVERSED IOTIFIED E	LC	F3AA	ѩ	-	4159
COMBINING CYRILLIC LETTER REVERSED IOTIFIED E	SC	F3AB	Ѩ̇	-	4142
CYRILLIC CAPITAL LETTER REARRANGED IOTIFIED E	UC	F3AC	Ѫ	-	-
CYRILLIC SMALL LETTER REARRANGED IOTIFIED E	LC	F3AD	ѫ	-	-
COMBINING CYRILLIC LETTER REARRANGED IOTIFIED E	SC	F3AE	Ѫ̇	-	-
COMBINING CYRILLIC LETTER CLOSED LITTLE YUS	SC	F3AF	Ѭ̇	86	4232
CYRILLIC CAPITAL LETTER CLOSED LITTLE YUS WITH VERTICAL STROKE	UC	F3B0	Ѭ	87	4259

*Proposal for a unified encoding of Early Cyrillic glyphs
in the Unicode Private Use Area*

CYRILLIC SMALL LETTER CLOSED LITTLE YUS WITH VERTICAL STROKE	LC	F3B1	Ɑ	87	4274
COMBINING CYRILLIC LETTER CLOSED LITTLE YUS WITH VERTICAL STROKE	SC	F3B2	Ɱ	87	4262
CYRILLIC CAPITAL LETTER LITTLE YUS WITH TRIANGULAR FOOT	UC	F3B3	Ɐ	-	4186
CYRILLIC SMALL LETTER LITTLE YUS WITH TRIANGULAR FOOT	LC	F3B4	Ɒ	-	4208
COMBINING CYRILLIC LETTER LITTLE YUS WITH TRIANGULAR FOOT	SC	F3B5	ⱱ	-	-
CYRILLIC CAPITAL LETTER THREE-STROKE LITTLE YUS	UC	F3B6	Ⱳ	-	4188
CYRILLIC SMALL LETTER THREE-STROKE LITTLE YUS	LC	F3B7	ⱳ	-	4210
COMBINING CYRILLIC LETTER THREE-STROKE LITTLE YUS	SC	F3B8	ⱴ	-	-
COMBINING CYRILLIC LETTER YA	SC	F3B9	Ⱶ	88	4292
COMBINING CYRILLIC LETTER IOTIFIED LITTLE YUS	SC	F3BA	ⱶ	92	4426
COMBINING CYRILLIC LETTER IOTIFIED CLOSED LITTLE YUS	SC	F3BB	ⱷ	93	4460
CYRILLIC CAPITAL LETTER IOTIFIED CLOSED LITTLE YUS WITH VERTICAL STROKE	UC	F3BC	ⱸ	94	4487
CYRILLIC SMALL LETTER IOTIFIED CLOSED LITTLE YUS WITH VERTICAL STROKE	LC	F3BD	ⱹ	94	4502
COMBINING CYRILLIC LETTER IOTIFIED CLOSED LITTLE YUS WITH VERTICAL STROKE	SC	F3BE	ⱺ	94	4490
CYRILLIC CAPITAL LETTER HOLLOW BIG YUS	UC	F3BF	ⱻ	-	-
CYRILLIC SMALL LETTER HOLLOW BIG YUS	LC	F3C0	ⱼ	-	-
COMBINING CYRILLIC LETTER HOLLOW BIG YUS	SC	F3C1	ⱽ	-	-
CYRILLIC CAPITAL LETTER CLOSED HOLLOW BIG YUS	UC	F3C2	Ȿ	-	-
CYRILLIC SMALL LETTER CLOSED HOLLOW BIG YUS	LC	F3C3	Ɀ	-	-
COMBINING CYRILLIC LETTER CLOSED HOLLOW BIG YUS	SC	F3C4	Ɀ̇	-	-
COMBINING CYRILLIC LETTER BLENDED YUS	SC	F3C5	Ɀ̈	90	4360
COMBINING CYRILLIC LETTER YN	SC	F3C6	Ɀ̉	91	4392
CYRILLIC CAPITAL LETTER REVERSED IOTIFIED BIG YUS	UC	F3C7	Ɀ̊	-	4518
CYRILLIC SMALL LETTER REVERSED IOTIFIED BIG YUS	LC	F3C8	Ɀ̋	-	4538
COMBINING CYRILLIC LETTER REVERSED IOTIFIED BIG YUS	SC	F3C9	Ɀ̌	-	-

COMBINING CYRILLIC LETTER KSI	SC	F3CA	Ѹ	96	4563
COMBINING CYRILLIC LETTER PSI	SC	F3CB	ѹ	97	4608
CYRILLIC CAPITAL LETTER IK	UC	F3CC	Ѻ	99	4679
CYRILLIC SMALL LETTER IK	LC	F3CD	ѻ	99	4695
COMBINING CYRILLIC LETTER IK	SC	F3CE	Ѽ	99	4683
COMBINING CYRILLIC LETTER IZHITSA	SC	F3CF	ѽ	100	4718
COMBINING CYRILLIC LETTER IZHITSA WITH DOUBLE GRAVE AKCENT	SC	F3D0	Ѿ	-	-

<i>Modifier Characters</i>					
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
COMBINING CYRILLIC INVERTED TITLO	MC	EF60	Ҁ	-	0436
COMBINING CYRILLIC ANGULAR TITLO	MC	EF61	ҁ	-	0440
COMBINING CYRILLIC REVERSED PAYEROK	MC	EF62	҂	4	0254
COMBINING INVERTED KREMASTY	MC	EF63	҃	31	-
COMBINING REVERSED KREMASTY	MC	EF64	҄	32	-
COMBINING CYRILLIC DOUBLE PSILI PNEUMATA	MC	EF65	҅	18	-
COMBINING CYRILLIC DOUBLE DASIA PNEUMATA	MC	EF66	҆	25	-
DAMAGE SIGN	MC	EF67	҇	-	-
LOST SIGN	MC	EF68	҈	-	-

<i>Punctuation Characters</i>					
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>
REVERSE TILDE	PC	EF70	Ҁ	-	0021
SLAVONIC KOPYE	PC	EF71	ҁ	8	0033
MIDDLE COMMA	PC	EF72	҂	2	-
REVERSED THREE DOT PUNCTUATION	PC	EF73	҃	13	0040
LEFT MARGINALIA	PC	EF74	҄	22	0221
RIGHT MARGINALIA	PC	EF75	҅	23	0222
TELIA	PC	EF76	҆	6	0146

About the authors ...

Victor A. Baranov is Professor and Chair of the Department of Linguistics at the Izhevsk State Technical University. His main fields of specialization are the history of Russian language, computational and applied linguistics, full-text databases, and research and publication of ancient Slavonic manuscripts. He is member of the International Commission on the Computer-Supported Processing of Medieval Slavonic Manuscripts and Early Printed Books to the International Committee of Slavists, leader of the Manuscript project (cf. http://manuscripts.ru/index_en.html).

David J. Birnbaum is Professor and Chair of the Department of Slavic Languages and Literatures at the University of Pittsburgh, where his research concentrates on applications of information technology to the study of medieval Slavic manuscripts. He is a member and past president of the International Commission on the Computer-Supported Processing of Medieval Slavonic Manuscripts and Early Printed Books to the International Committee of Slavists and has served on the board of directors of the Association for Computers and the Humanities (ACH) and the *Technical Council* of the Text Encoding Initiative (TEI). He is also one of the founders of the Repertorium of Old Bulgarian Literature and Letters project.

Ralph Cleminson is a specialist in the paleography and codicology of Cyrillic manuscripts and earlyprinted books, and is currently working on textual criticism of the Slavonic version of the New Testament. Like the other authors of this paper, he has been instrumental in expanding the Cyrillic and Glagolitic repertory of Unicode, and is currently president of the Commission on the Computer-Supported Processing of Medieval Slavonic Manuscripts and Early Printed Books to the International Committee of Slavists.

Heinz Miklas is Professor of Slavic Philology at the Institute of Slavic studies of Vienna University and Foreign member of the Bulgarian Academy of Sciences. Presently, his main research interests are focused on Old Slavonic (Cyrillic and Glagolitic) paleography and graphemics. He is currently engaged in the analysis and edition of the newly found Glagolitic manuscripts on Mt. Sinai (cf. <http://www.caa.tuwien.ac.at/cvl/research/sinai/>).

Achim Rabus received his PhD in 2008. Currently, he is an Assistant Professor of Slavic Linguistics at the University of Freiburg, Germany. His research interests include contact linguistics, Church Slavonic as well as Ruthenian language and culture, and computer-assisted processing of medieval Slavic manuscripts.