

Proposal to Encode Additional Cyrillic Characters used in Early Church Slavonic Printed Books

Aleksandr Andreev^{*}

Yuri Shardt

Nikita Simmons

PONOMAR PROJECT

1 Introduction

Church Slavonic (or Church Slavic, ISO 639-2 code *cu*) is a highly codified, living, literary language used by the Slavs. Presently, various recensions of Church Slavonic are used by Slavic Orthodox Churches, such as the Russian Orthodox Church, and by Slavic Byzantine-Rite Catholic Churches as primary or secondary liturgical languages. Historically, the language was used not only for liturgical texts and religious literature but also for secular academic literature, such as grammars, lexicons, and even astronomical treatises, prior to the codification of literary recensions of vernacular Slavic languages. Church Slavonic has been written in two scripts – Cyrillic and the earlier Glagolitic. This proposal focuses on Church Slavonic texts printed in the Cyrillic alphabet only.

In this document, the authors identify nine Cyrillic characters used in early Church Slavonic printed books published between 1500 and 1700. The characters can be classified as being of three types: characters used in historically important literary texts (such as the Bible printed by Francysk Skaryna, the Ostrog Bible printed by Ivan Fedorov, or the Trebnik compiled by Metropolitan Peter Mogila). These texts are of immense value to academics studying the history of Slavic languages, the Cyrillic printed tradition, and the Byzantine Rite. The second class constitutes characters that are in use in liturgical texts published today by the Russian Old Ritualist community. Old Ritualists rejected the linguistic and liturgical reforms carried out in Russia under Patriarch Nikon and his successors (1653-1700) and continue to use liturgical books that imitate the earlier printed tradition. The third class constitutes characters that are in use in modern liturgical texts printed by the mainline Russian Orthodox Church, especially those published prior to the Russian Revolution in Kiev at the Lavra of the Caves and their reprints in the Russian diaspora. Characters of the third category are usually used contextually, but may sometimes occur outside of context.

^{*}Corresponding author: aleksandr.andreev@gmail.com.

Table 1: Table of Proposed Characters

Codept	Glyph	Name
U+1C80	Ѣ	CYRILLIC SMALL LETTER ROUNDED VE
U+1C81	ѣ	CYRILLIC SMALL LETTER LONG-LEGGED DE
U+1C82	Ѧ	CYRILLIC SMALL LETTER NARROW O
U+1C83	ѧ	CYRILLIC SMALL LETTER WIDE ES
U+1C84	Ѩ	CYRILLIC SMALL LETTER TALL TE
U+1C85	ѩ	CYRILLIC SMALL LETTER THREE-LEGGED TE
U+1C86	Ѫ	CYRILLIC SMALL LETTER TALL HARD SIGN
U+1C87	ѫ	CYRILLIC SMALL LETTER TALL YAT
U+1C88	Ѭ	CYRILLIC SMALL LETTER UNBLENDED UK

The present document proposes to encode these nine characters in a new block titled Cyrillic Extended-C. The characters in question are listed in Table 1. The next section explains the usage of each character in detail; Section 3 presents a justification for their encoding; and Section 4 discusses some technical matters.

2 Proposed Characters

2.1 Cyrillic Rounded Ve

This character is a variant form of the Cyrillic Letter Ve, known as the “rounded Ve” for its characteristic shape. The base form is encoded at U+0432. The Rounded Ve appears in incunabula of a West Slavic provenance as well as in later Poluustav texts printed in the Polish-Lithuanian Commonwealth. In Figure 1 we present an example from the Bible of Francysk Skaryna, printed in Prague circa 1519. In this particular example, the rounded form is used whenever the letter Ve does not take a diacritical mark (combining letter, titlo, or payerok) and the base form is used whenever the letter Ve occurs with a combining mark. However, observe that when the letter Ve occurs under a titlo (as the numeral two), the rounded form is used. Thus, the rounded form cannot be specified algorithmically. In addition, the usage observed in Skaryna’s edition of Exodus does not hold elsewhere in his Bible. Sometimes Skaryna uses the Rounded Ve in certain words and the base form in others while in other places, Skaryna’s usage of the two characters for Ve is haphazard, as can be seen from Figure 2. Figure 3 provides the list of characters used by Francysk Skaryna in his Bible.

In addition to its use in the Bible of Skaryna, the rounded form of Ve also occurs in other incunabula¹ printed in Cetinje, Kraków, Prague, Venice, and Vilnius. An example from an Horologion is presented in Figure 4. Critical studies of these sources have yet to be undertaken, and so the extent to which any given rules are followed is yet to be established. However, based on our

¹An incunabulum (or incunable) is a book, pamphlet, or broadside printed before the year 1501 in Europe. We use the term incunabulum more broadly to apply to the books printed in South and West Slavia up to the mid-16th Century since they form a distinct and unique printing tradition.

study of the sources, it is clear that the usage of this character cannot be predicted algorithmically in many cases. This character is not used in modern liturgical texts; it is only used historically in important printed works.

2.2 Long-Legged De

This character is a variant of the Cyrillic letter De, known as the “Long-legged De”. The base form is encoded at U+0434. In the manuscript tradition of the Poluustav era, the base form was written in the initial position while the long-legged form was written in the medial or final positions (Karsky, 1979, p. 186). This convention was carried over to the printed texts published in the Polish-Lithuanian Commonwealth, and can be observed, for example, in the printed text of the Statutes of Lithuania, presented in Figure 5. However, unlike scribes, typesetters began to use both forms indiscriminately; this was particularly the case in Moscow, where the base form was used when the amount of vertical space between lines of text was limited or a collision needed to be avoided with diacritical marks on the line below. In general, studying the usage of the long-legged form vs. the base form reveals important details about the evolution of the print tradition and mutual influences of typographers on one another.

Figure 6 presents the typical usage of both the base form and the long-legged form in a text of Muscovite provenance. The usage of the two characters is completely haphazard and cannot be predicted algorithmically. This usage can still be observed today in texts published by Old Ritualists, who have maintained a print tradition that mimics the older Muscovite type forms. Eventually, the long-legged form completely fell out of use in Muscovite typography and, with rare exceptions, it does not occur in Synodal Church Slavonic texts of a Muscovite origin.

However, the long-legged form continued to be used extensively in Synodal editions of a Kievan provenance. Figure 7 presents an example from a book published in Kiev where it can be clearly observed that the base form is used in initial position and the long-legged form is used in medial position, in keeping with earlier rules of usage. However, whenever the letter De occurs as part of a numeral, only the base form is used. Figure 8 presents a calendrical chart out of the same book; in this figure, it can be observed that, when it occurs as part of a numeral, the letter De is encountered only in the base glyph form, regardless of position. Since in charts of numbers, the titlo used to indicate that the letters form a numeral is often omitted (as is the case in this example), it is impossible to predict algorithmically that the group of letters constitutes a numeral and thus impossible to use contextual rules to select the character.

In conclusion, the long-legged De is a character that is encountered in all three usages: in historical texts, in liturgical texts of the Russian Old Ritualist communities, and in modern liturgical texts of a Kievan provenance.

2.3 Narrow O

This is a variant of the Cyrillic letter O, known as the “Narrow O”. The base form has been encoded at U+043E. This narrow form is widely used in Slavonic typography of all recensions. In the earliest Poluustav printed texts, rules governing the usage of the Narrow O were not fixed, and

so this character may be found both in the medial and the final positions and may be either accented or unaccented. This can be observed in Figure 9. The character was used both in texts printed in Moscow and in the Polish-Lithuanian Commonwealth, as can be seen in Figure 12.

In later printed editions, the usage stabilized and the narrow form came to be used whenever the letter O does not take an accent, while the base form was used in the accented position. This practice continues to be maintained in modern texts printed by the Russian Old Ritualists, though this usage was not always adhered to strictly, as can be observed from Figure 10. In addition, the typical usage of these characters may be reversed in compound words. Thus, in Figure 11, we observe the narrow form used in the unaccented position everywhere except as the last character of the first part of a compound word; the base form is used in that position for semantic purposes.

In addition to the base form, Unicode includes the wide form of the Letter O, called “Round Omega” (U+047B). This wide form originates in Moscow and is used in modern Church Slavonic orthography in a very specific circumstance: only in the initial position, for example, in the word *О҃и҃х* (*father*), or, in the medial position, as the initial letter of a stem in a compound, as in the word *п҃ро҃и҃х* (*forefather, ancestor*). Since it has a specific grammatical function – to indicate the first letter of a root that starts with *o* – this form should not be used to encode the base form of the Letter O (U+043E). We can observe from Figure 10 that all three forms of the Letter O (U+043E, U+047B and the Narrow O) may occur in a typeface and all may be either accented or unaccented. It follows that it would not be correct to use the “Round Omega” to encode the base form of the Cyrillic Letter O and to use the codepoint of the Cyrillic Letter O to encode the narrow form.

In modern Synodal typography, the Narrow O is encountered extremely rarely, only as an apparent space-saving device. In addition, both in Poluustav and in Synodal recension texts, the narrow form of O occurs as the first glyph of the digraph letter *oy*. In fact, writing *oy* instead of *oy* is generally not correct, though it does occur in some publications, notably in the 1619 *Grammar* of Meletius Smotrytsky. Unicode had initially encoded the digraph Uk as a standalone character (U+0479). However, the typographic tradition strongly suggests that it is properly treated as two glyphs; for example, when in a text the initial letter of a paragraph is set in red type, it is typical for only the *o* glyph to be set in red and not the entire *oy* digraph. Likewise, the capitalized form of the digraph may be either *Oy* or *Oy*, depending on the context. Thus, the codepoints U+0478 and U+0479 should not be used and the digraph *oy* is properly encoded as either U+043E U+0443 or as the proposed U+1C82 U+0443, depending on the usage.

2.4 Wide Es

This is a variant form of the Cyrillic Small Letter Es, known as the “Wide Es”. The base form has been encoded at U+0441. This form is only encountered in initial position and only in texts of a Kievan provenance. In Kievan texts of the Synodal recension – that is, modern liturgical texts of the Russian Orthodox Church – this variant form is used in words that refer to the Divinity but are not divine names (*nomina sacra*). This can be clearly seen from Figure 14. Observe that the variant form (boxed in red) is used as the initial letter of the word *с҃вѣтъ* (*light*) when it refers to Christ (“light of the Gentiles”, an allusion to Luke 2:32). On the same page, we observe the base form of the letter used in the same word *с҃вѣтъ* (*light*) when it refers to a saint (“light upon a candlestick”, an allusion to Matthew 5:15). Thus, the variant is used in the first example simply to

distinguish that the word light in this context refers to Christ. Observe also that the variant form is **not** a capital form of the letter Cyrillic Es, since the capital form may also be seen on this page in the word **СѸМЕѠНЪ** (Symeon), a proper name, and in the word **СѸИѠ** (Savior), a *nomen sacrum*, both boxed in black.

In earlier printed texts of the Poluustav era, the typographical and orthographic rules were less rigid, but the same general pattern of usage may be observed. Figure 13 presents an example from the *Trebnik* (Euchologion) compiled by Metropolitan Peter (Mogila), a monumental 17th Century text that is still important both as a practical reference for clergy and as a fundamental primary source for the study of the development of Eastern Orthodox ritual. In this text, both forms of the letter Es are encountered (as well as the capitalized form), though the pattern of usage is less clear. The base form appears to be used in conjunctions and other less important words while the Wide Es is used for nouns. With regular frequency, the demonstrative pronoun **ѠН** (this one) and its oblique case forms are written with the base form when they refer to an object or concept, and written with the wide form when they refer to a person.

Finally, the Wide Es may also be encountered in modern Church Slavonic texts in marginal notes as an abbreviation for the word **ѠНЪ** (*verse*), accompanying verse numbering in Psalms or Old Testament Lessons reprinted in liturgical books. An example is provided in Figure 15.

2.5 Tall Te

This is a variant form of the Cyrillic Te, known as the “Tall Te”. The base form is encoded at U+0442. This character commonly occurs in Poluustav printed editions, where it reflects archaic spelling conventions inherited from the manuscript tradition. The typical usage is demonstrated in Figure 16. Often the character functions as a space-saving device, since liturgical texts are commonly laid out in justified alignment and in the earliest printed Church Slavonic texts hyphenation (the transfer of a portion of a word to a new line) was avoided out of a desire for theologically precise language. When texts are typeset without any hyphenation, if the amount of inter-word spacing cannot be further reduced, the Tall Te character can be used.

In storing digital versions of these early Slavonic printed texts, it is important to preserve the use of this character for two reasons. On the one hand, scholars need to use digital methods to study and analyze the typographic, orthographic, and linguistic conventions used by early typographers. On the other hand, in producing either reprints of older texts (such as the *Trebnik* of Peter Mogila) or new texts for liturgical use, there is often a need to reproduce early typographic conventions exactly; for example, many Old Ritualist texts are still printed without hyphenation, and thus the Tall Te character is still used. We discuss this at length in a following section.

We note also that in some instances, adjoining letters Te and Ve are written in Slavonic texts as the ligature **ѠѢ**; in these cases, the Tall Te character forms the first component of the Te-Ve ligature, as can be seen in Figure 17. The Unicode standard does not encode ligatures (they may be handled via the use of U+200D ZERO WIDTH JOINER); however, instances may arise where the two components need to be displayed independently (as in **Ѡ Ѣ**), for example in producing software manuals or instructional materials on Slavonic typography.

2.6 Three-Legged Te

This is another variant of the Cyrillic Letter Te (U+0442), called the “Three-Legged Te” and written with all three vertical strokes touching the baseline. According to Karsky (1979, p. 198), in the 15th Century, this became the most prevalent form of the letter Te in Church Slavonic manuscripts. What later becomes the standard form (U+0442) is originally used in the manuscript tradition as a space-saving device. This is demonstrated in Figure 18. In printed editions of Church Slavonic texts, the standard form (U+0442) begins to dominate, and the Three-Legged Te gradually drops out of usage, especially in Moscow. Some editions, particularly those published in the Polish-Lithuanian Commonwealth, however, use both forms interchangeably, out of a desire to imitate the manuscript tradition. We demonstrate an example of this usage in Figure 19. Observe that the Three-Legged Te is used everywhere in initial position while the standard form (U+0442) is used in medial position, reflecting the earlier manuscript orthography. However, in words of a Greek origin, the standard form is used in initial position, undoubtedly because it imitates the Greek letter tau (τ). Similar usage of the standard form in words of a Greek origin is demonstrated in Figure 20. As can be observed, the usage of these two characters cannot be predicted algorithmically, and so both forms must be encoded.

2.7 Tall Hard Sign

This character is a variant of the Cyrillic Hard Sign, called the “Tall Hard Sign.” The base form has been encoded at U+044A. The Tall Hard Sign appears both in books printed in Moscow and in the Polish-Lithuanian Commonwealth. In Muscovite publications its usage appears to be haphazard, as can be seen in Figure 21. However, the character plays a semantic function in some texts printed in the Commonwealth, particularly in the Ostrog Bible. In Old Church Slavonic, the Hard Sign was pronounced as a semivowel. However, the semivowel pronunciation was eventually lost and the Hard Sign came to be pronounced the same way as the Letter O (for a discussion of this see Uspensky (1987, p. 138ff)). This assimilation came to be reflected in the orthography, where the Hard Sign in medial position came to be replaced by the Narrow O. However, in some texts, out of a desire to convey older, archaic orthographic conventions, the Hard Sign spelling was retained. In Figure 22 and Figure Figure 23, we observe the Tall Hard Sign used in the Ostrog Bible in the medial position where it is to be pronounced the same way as the Narrow O.

2.8 Tall Yat

This character is a variant form of the Cyrillic Letter Yat; the standard form is encoded at U+0463. The Tall Yat appears in some printed publications in word roots that contain the letter ѣ as a reflection of archaic spelling norms inherited from the manuscript tradition. We present usage from a Gospel Book printed in Moscow in Figure 24 and from the Ostrog Bible in Figure 25.

2.9 Unblended Uk

In Synodal Church Slavonic, the orthography has been standardized and the digraph ou (encoded as U+043E U+0443) always occurs in the beginning of words while the monograph form u (U+A64B) occurs in medial or final position (and never in initial position). However, such standardization is not the case in earlier recensions, including printed Poluustav Church Slavonic texts, where the monograph form and the digraph form of the letter are used more or less interchangeably. In addition, in some Poluustav editions, particularly those printed in the Polish-Lithuanian Commonwealth, we encounter another form of the letter Uk, which is called the “Unblended Uk.” The form is “unblended” because the o and the y portions of the character have not been assimilated (blended) as in u , but rather each retains its characteristic shape in the grapheme Uk . The form appears in certain words in initial position, as can be seen in Figure 26 as well as in the medial and final positions, as can be seen in Figure 27. Note that in all of these examples, the base form u and the digraph form ou also occur. Since most of the texts in which the Unblended Uk occurs were printed in modern-day Belarus and Lithuania, it is possible that this form is a graphemic precursor to the modern character y used in Belorussian.

3 Justification

The characters in this proposal are used in two settings: by academics in the reproduction of historical texts or quotation of historical texts in their research and in the production of modern liturgical texts used by the Russian Old Ritualist communities and (in some instances) by the mainline Russian Orthodox Church. As such, the justification for the encoding of these character is twofold.

The texts cited in this proposal are of an immense historical value. For example, the Ostrog Bible is the first complete Bible printed in the Church Slavonic language. Its publication had an important influence on the literary culture of the Eastern Orthodox communities in the Polish-Lithuanian Commonwealth. Subsequently, it was reprinted in Moscow (with minor orthographic variation) as the Moscow Bible of 1663, which continues to be in use today by the Russian Old Ritualist communities. Despite its historical value, critical studies of the text of the Ostrog Bible have yet to be undertaken (Nemirovsky, 2003). The Bible of Francysk Skaryna is of considerable linguistic importance because its language is closer to the contemporary vernacular spoken in the Polish-Lithuanian Commonwealth than to literary Church Slavonic; thus, it has been identified as the first attempt to translate the Bible into a vernacular Slavic language (the ancestor to modern Ukrainian and Belorussian). In addition, Skaryna published commentaries to his Bible, which were viewed as heretical by the Eastern Orthodox hierarchy but heavily influenced Protestant communities in the Commonwealth (Podokshin, 1981, p. 8ff.). Despite its important role in the history of Slavic languages and East-Slavic religious thought, it, too, is yet to be critically studied. Much more can be said also of the importance of the Trebnik compiled by Metropolitan Peter (Mogila), the early printed texts produced by the Anonymous Press in Moscow (the first attempts at book printing in Moscow), and the other printed editions cited.

While manuscripts studied by palæographers exhibit considerable graphemic variation due to the different usages of scribes, printed texts show much more uniformity. Thus, scholars who study these texts are able to correctly identify the *character repertoire* used in a given text. This repertoire includes characters that may no longer be used in modern texts printed in Cyrillic, but that were used historically. The study of the character repertoire, including archaic characters, allows scholars to draw sound, well-grounded conclusions about the Cyrillic writing system and the early printed texts. Such study requires scholars to rely on *bona fide* transcriptions of a text rather than simple transliterations into modern characters (“normalized editions”). Encoding archaic characters makes such transcriptions technically possible, since the full character repertoire becomes available in computer fonts. Furthermore, encoding these characters allows scholars to use computer methods to study these texts. Finally, the encoding of these characters facilitates the creation of authentic digital copies of historical texts, allowing scholars to faithfully preserve the literary heritage of Eastern Europe.

The study of these historical texts at the graphemic level has a further importance because of the history of the Church Slavonic literary language and its reforms. Liturgical and linguistic reforms of the Church Slavonic liturgical texts were carried out in the Polish-Lithuanian Commonwealth by Metropolitan Peter (Mogila) around 1635-1645, and included the publication of new prayer books as well as new editions of the *Sluzhebnyk* and *Trebnik* (Euchologion) (Meyendorff, 1985). Liturgical and linguistic reforms were later carried out in Moscow by Patriarch Nikon and his successors around 1653-1698. The Nikonian reforms were heavily influenced by the earlier reforms of Metropolitan Peter, and scribes and scholars from Kiev were active in Moscow (Uspensky, 1975). The orthographic usage of the various printed texts during these reforms allows scholars to shed considerable light on the degree of influence. Furthermore, the reforms were motivated at the official level by a desire to “correct the errors” that had allegedly crept into liturgical books. These errors or perceived errors – including irregular orthography, nonce glyphs, and other graphemic variation – have to be studied critically. All in all, the encoding of these characters would help scholars of Slavic languages and literature and of the Byzantine Rite.

In addition to their use in publications of interest to scholars, these characters also occur in liturgical texts used by the Russian Old Ritualist communities. Old Ritualists rejected the reforms of Patriarch Nikon and his successors and continued to use the earlier printed books. The publication of books by Old Ritualists was forbidden by law in imperial Russia. However, in 1798 a number of Old Ritualist communities were admitted into communion with the mainline Russian Orthodox Church but allowed to maintain and reprint the pre-reformed books in a movement called *Yedinoverie* (the “one-faith movement”). The books published by the *Yedinovertsy* faithfully represented all elements of the pre-Nikonian orthography, including the use of the archaic characters. In 1905, all restrictions against Old Ritualists in Russia were lifted and a number of editions of liturgical books were printed by Old Ritualist communities prior to the Russian Revolution. Nowadays, the encoding of these additional Cyrillic characters is necessary to facilitate the production of Old Ritualist and *Yedinovertsy* liturgical books using modern computer techniques.

While historical texts should not be normalized because scholars wish to study their graphemic content, the graphemic content of liturgical materials should be preserved because of a partic-

ular way in which Old Ritualists approach the sacred text. To explain this approach, we must introduce some terminology from the semiotics literature. A writing system with “conventional semiotics” is one where the symbols themselves carry no meaning but rather are assigned meaning by convention between the two parties to the written communication. Thus, in writing “the”, neither the individual symbols nor the sequence in its entirety have any meaning in and of themselves; rather, it has been agreed by society that this is the way to record the definite article in English (a “convention”) and thus in the context of English, the sequence of characters obtains a conventional meaning. English – and most modern languages – are treated as having conventional semiotics. On the other hand, as Uspensky (1994) has argued, Church Slavonic in mediæval Russia was interpreted as having “non-conventional semiotics;” that is, the written symbol was not interpreted as a conventional representation of some subject but was treated as being isomorphic to the subject itself or as a part of the subject. The sacred text written down with these symbols was perceived in a non-conventional manner, and the correctness of religious rites was seen to be dependent not only upon the content of the sacred text but also upon the faithfulness of the symbolic representation. An aberration in the symbolic representation (for example, a different orthography), even if the underlying meaning remains the same, could be interpreted as blasphemous or heretical by the user of the written text.

This approach to written text is not unique to Church Slavonic but is shared by many other societies that use languages of a sacred function (for example, Koranic Arabic). In the context of Cyrillic, Uspensky (1994) argues that one of the reasons behind the Old Ritualist schism was precisely the issue of semiotics. In many situations, the new (“reformed”) texts were different from the old texts only in orthography and not in meaning; but the new orthography was interpreted by some – those who were not ready for a conventional interpretation of the text’s semiotics – as heretical and sufficient grounds to reject the new text. Thus, many Old Ritualists continue to interpret the text in a non-conventional manner. For this reason, computer specialists must strive to represent Church Slavonic texts in a manner that is faithful to the written tradition, keeping in mind that the recipient of the electronic communication may have a particularly devout attitude toward the way the written text is presented, not just to the message it conveys.

3.1 Urgency

The Ponomar Project (<http://www.ponomar.net/>) is making Old Ritualist liturgical texts available in an electronic format (as HTML). Scholars are also working on the digital preservation of important historical editions. For example, the Ponomar Project is preparing for electronic publication the text of the Ostrog Bible. As long as text is being interchanged in a closed software-platform setting between a limited number of users, any encoding methodology (such as font mapping or the use of the Private Use Area) can be used. But since these texts are becoming available in electronic format online and are being indexed by major search engines, a standard encoding model is urgently needed. Thus, the authors request that this proposal be expedited.

4 Technical Issues

So far, all Cyrillic characters encoded in the Unicode standard have been encoded with both an upper-case and a lower-case form. The characters under consideration presently do not have extant uppercase forms. Given the general structure of the Cyrillic script, it would not be advisable to encode these characters without case mapping. In scholarly publications, for example, titles are often set in all caps and encoding a few characters that cannot be operated upon by capitalization routines would be at odds with existing implementations. However, we do not wish to construct non-existent Cyrillic graphemes. Rather than constructing such uppercase forms artificially, we propose that only the lower-case forms be encoded and that the archaic characters uppercase to the upper-case forms of the modern character. The casing operations will thus fold the archaic characters onto the modern characters. Precedent for this has been set with the encoding of U+03C2 GREEK SMALL LETTER FINAL SIGMA. In addition to alleviating the need for “inventing” upper-case forms for these characters, such an approach is also simpler from the standpoint of text processing because it allows for string manipulation and comparison by using simple case folding rather than collation tables. Thus, the fact that casing relationships for Cyrillic cease to be isomorphic in this case is a benefit rather than a drawback.

So that casing operations for Cyrillic characters take place entirely along the Basic Multilingual Plane (BMP), we propose that these characters be encoded in an empty column of the BMP.

Keeping in mind these recommendations, the following data are proposed for addition to `UnicodeData.txt`. The Appendix provides a proposed codechart.

```
1C80;CYRILLIC SMALL LETTER ROUNDED VE;L1;0;L;;;;;N;;;0412;;0412
1C81;CYRILLIC SMALL LETTER LONG-LEGGED DE;L1;0;L;;;;;N;;;0414;;0414
1C82;CYRILLIC SMALL LETTER NARROW O;L1;0;L;;;;;N;;;041E;;041E
1C83;CYRILLIC SMALL LETTER WIDE ES;L1;0;L;;;;;N;;;0421;;0421
1C84;CYRILLIC SMALL LETTER TALL TE;L1;0;L;;;;;N;;;0422;;0422
1C85;CYRILLIC SMALL LETTER THREE-LEGGED TE;L1;0;L;;;;;N;;;0422;;0422
1C86;CYRILLIC SMALL LETTER TALL HARD SIGN;L1;0;L;;;;;N;;;042A;;042A
1C87;CYRILLIC SMALL LETTER TALL YAT;L1;0;L;;;;;N;;;0462;;0462
1C88;CYRILLIC SMALL LETTER UNBLENDED UK;L1;0;L;;;;;N;;;A64A;;A64A
```

5 Illustrations

Figure 1: Cyrillic Small Letter Ve (U+0432; boxed in black) and Cyrillic Small Letter Rounded Ve (proposed at U+1C80; boxed in red). Source: Bible printed by Francysk Skaryna, Prague, circa 1519.

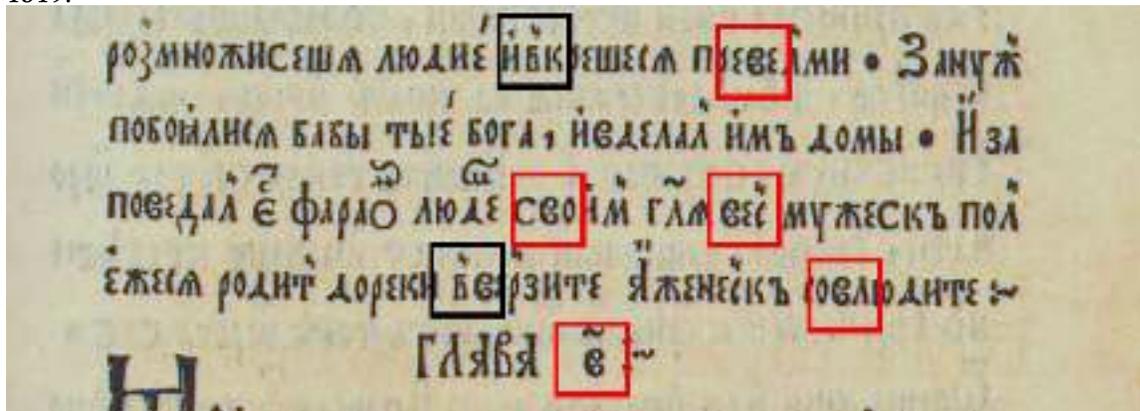


Figure 2: Cyrillic Small Letter Ve (U+0432; boxed in black) and Cyrillic Small Letter Rounded Ve (proposed at U+1C80; boxed in red). Source: Bible printed by Francysk Skaryna, *op. cit.*

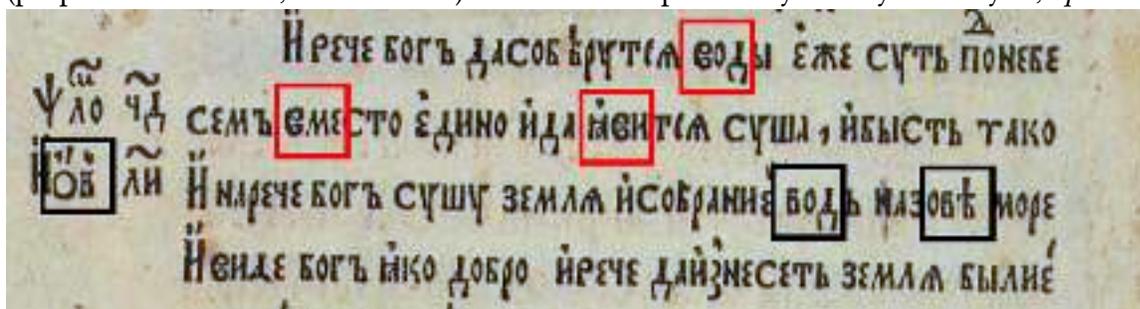


Figure 3: Characters used in the Bible of Francysk Skaryna. Note the use of two character for Ve: the standard form (U+0432; boxed in black) and the Rounded form (proposed at U+1C80; boxed in red).

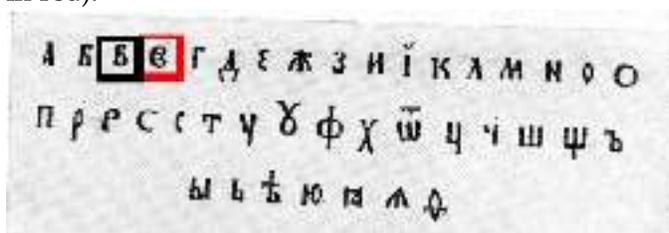


Figure 4: Cyrillic Small Letter Ve (U+0432; boxed in black) and Cyrillic Small Letter Rounded Ve (proposed at U+1C80; boxed in red). Source: Horologion (Book of the Hours), Vilnius, c. 1522.

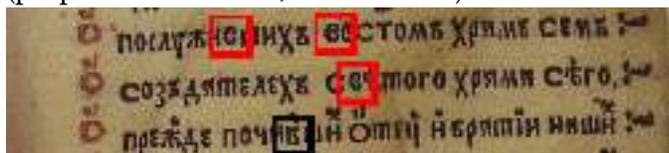


Figure 5: Cyrillic Small Letter De (U+0434; boxed in black) and Cyrillic Small Letter Long-legged De (proposed at U+1C81; boxed in red). Note that the long-legged form occurs in medial position while the base form occurs in initial position. However, the long-legged form is used in the initial position when it is preceded by a proclitic preposition, making algorithmic selection impossible. Source: title page of the Third Statutes of Lithuania, Vilnius: printing press of the Mamonichi Merchants, 1588.

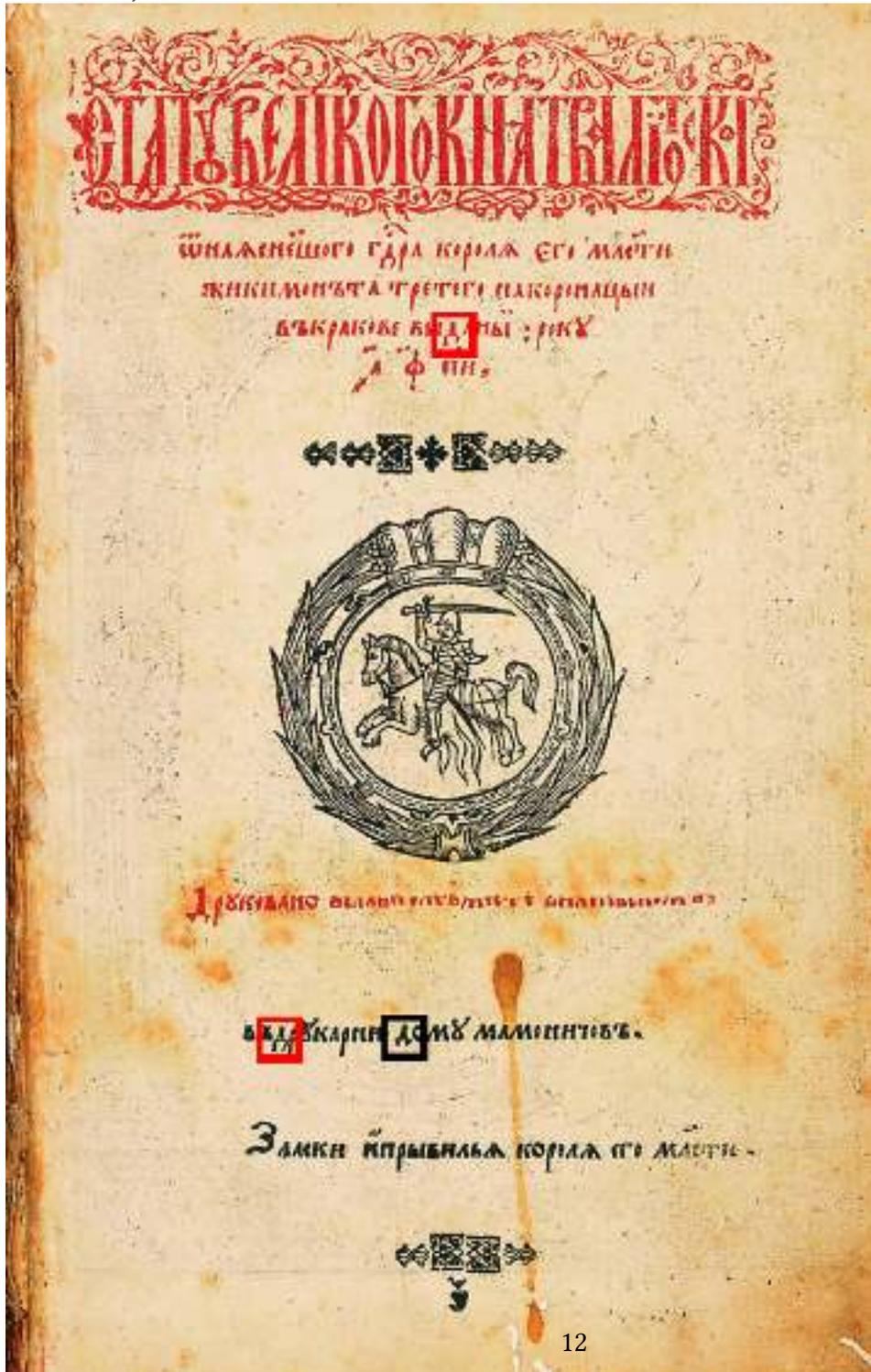


Figure 6: Cyrillic Small Letter De (U+0434; boxed in black) and Cyrillic Small Letter Long-legged De (proposed at U+1C81; boxed in red). Note that both forms can occur in initial or medial positions. Source: Apostolos, Moscow: Ivan Fedorov, 1564.

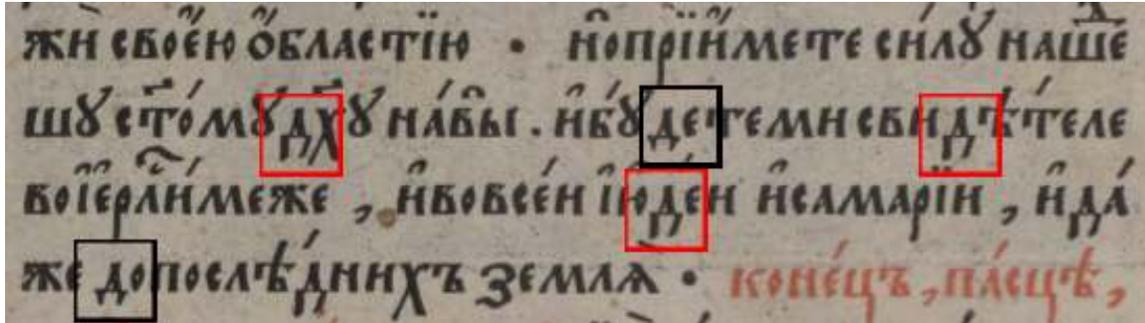


Figure 7: Cyrillic Small Letter De (U+0434; boxed in black) and Cyrillic Small Letter Long-legged De (proposed at U+1C81; boxed in red). Note that the long-legged form occurs in medial position while the base form occurs in initial position. Note also that the base form only is used for numerals. Source: Typicon, Kiev: Lavra of the Kiev Caves, 1893.

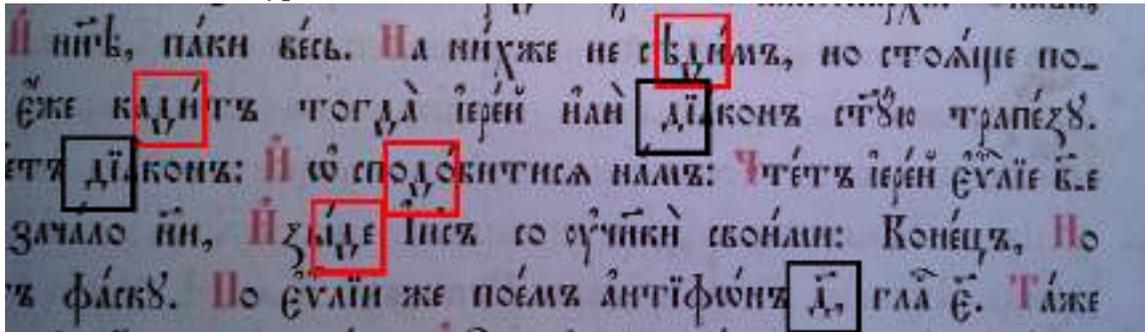


Figure 8: Cyrillic Small Letter De (U+0434; boxed in black) used for numerals. Source: Typicon, Kiev: Lavra of the Kiev Caves, 1893.

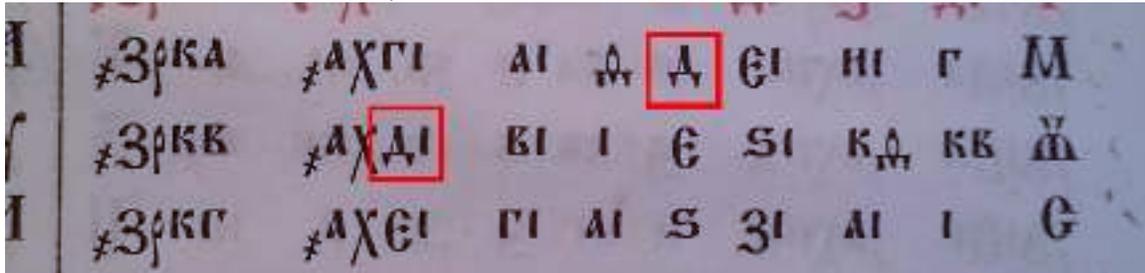


Figure 9: Cyrillic Small Letter O (U+043E; boxed in black) and Cyrillic Small Letter Narrow O (proposed at U+1C82; boxed in red). Note that both forms can occur with or without an accent. Source: *Oko Tserkovnoye* (Typicon), Moscow: official Printing Press, 1610.

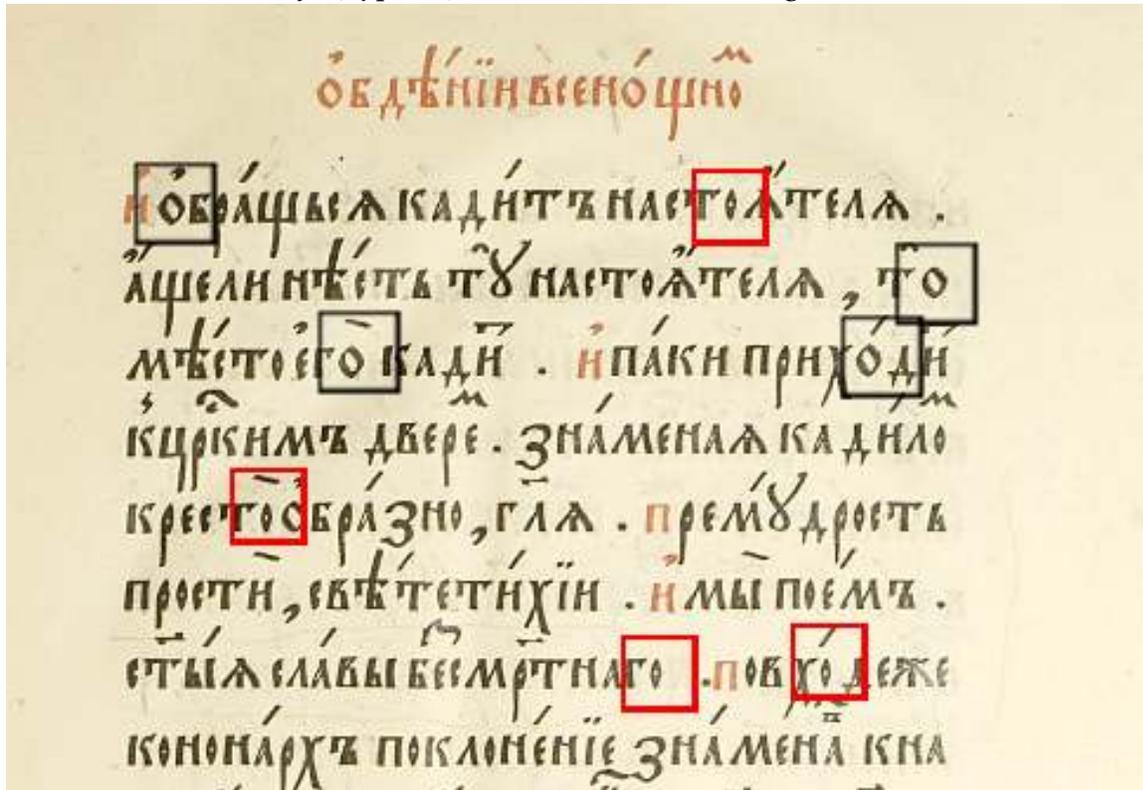


Figure 10: Note the use of three forms of the letter O – the typical Cyrillic Small Letter O (U+043E; boxed in black), the Cyrillic Letter Round Omega (U+047B; boxed in indigo) and the Cyrillic Letter Narrow O (proposed at U+1C82; boxed in red). Source: *Prologue*, Moscow: Printing Press of the Moscow Old Ritualist Community, 1915.



Figure 13: Cyrillic Small Letter Es (U+0441; boxed in black) and Cyrillic Small Letter Wide Es (proposed at U+1C83; boxed in red). Source: *Trebnik* (Euchologion) of Metropolitan Peter (Mogila), Kiev: Lavra of the Kiev Caves, 1646.

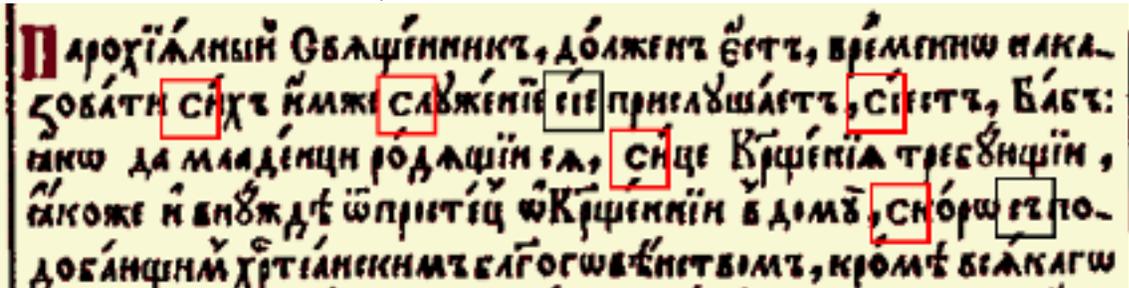


Figure 14: Cyrillic Small Letter Es (U+0441; boxed in blue) and Cyrillic Small Letter Wide Es (proposed at U+1C83; boxed in red). The capital form (U+0421) has been boxed in black. Source: *Menaion* for February, Kiev: Lavra of the Kiev Caves, 1893.

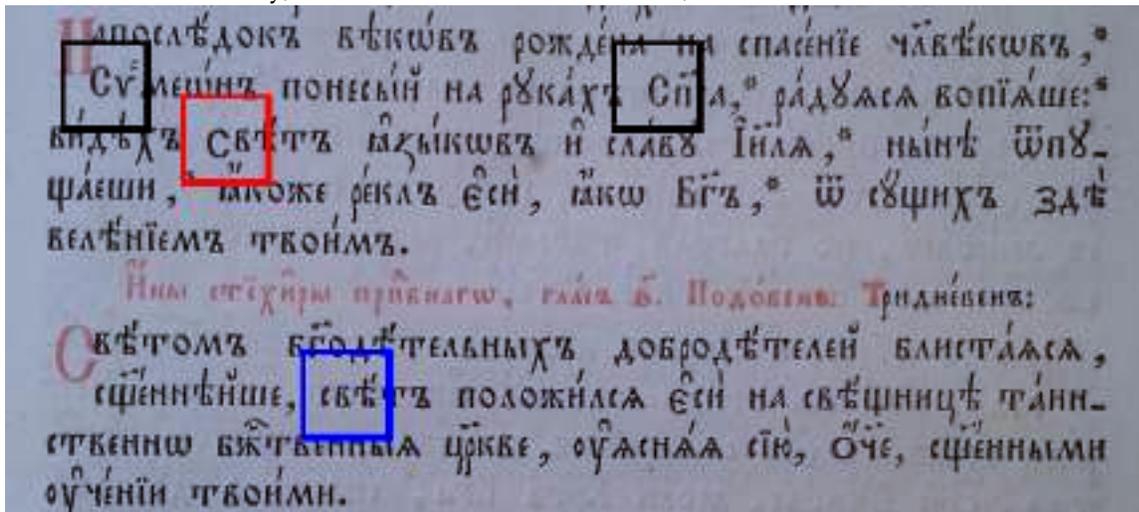


Figure 15: Cyrillic Small Letter Wide Es (proposed at U+1C83; boxed in red) used in marginal notes to indicate verse numbering. Source: Psalter, Jordanville, New York, USA: Holy Trinity Monastery, 1959.

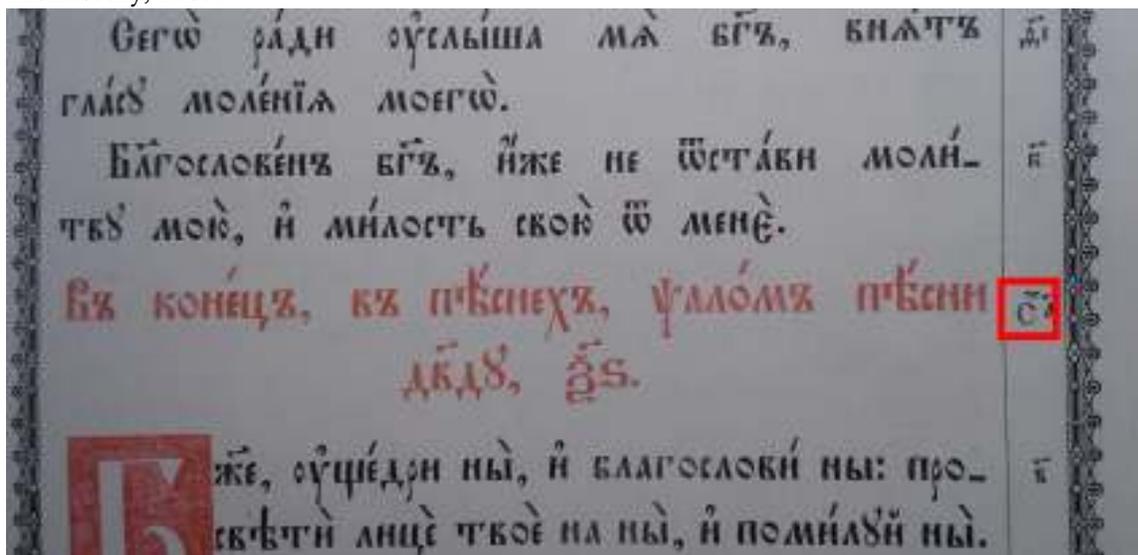


Figure 16: Cyrillic Small Letter Te (U+0442; boxed in black) and Cyrillic Small Letter Tall Te (proposed at U+1C84; boxed in red). The Tall Te appears to be used as a space-saving device. Source: *Trebnik* (Euchologion) of Metropolitan Peter (Mogila), *op. cit.*

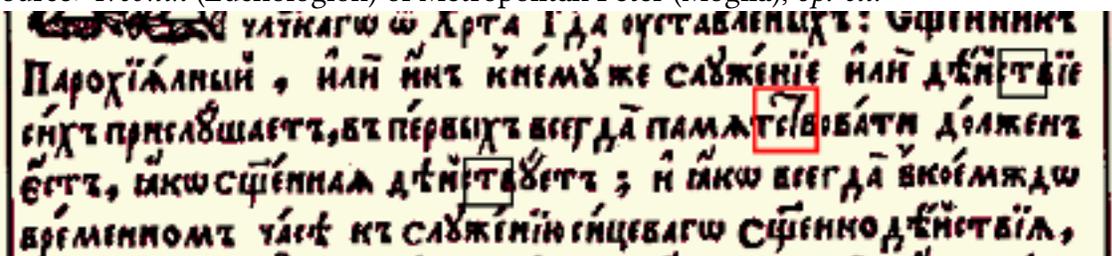


Figure 17: Cyrillic Ligature Te-Ve (boxed in red). The Cyrillic Small Letter Te occurs as the first component of this ligature. Source: *Trebnik* (Euchologion) of Metropolitan Peter (Mogila), *op. cit.*

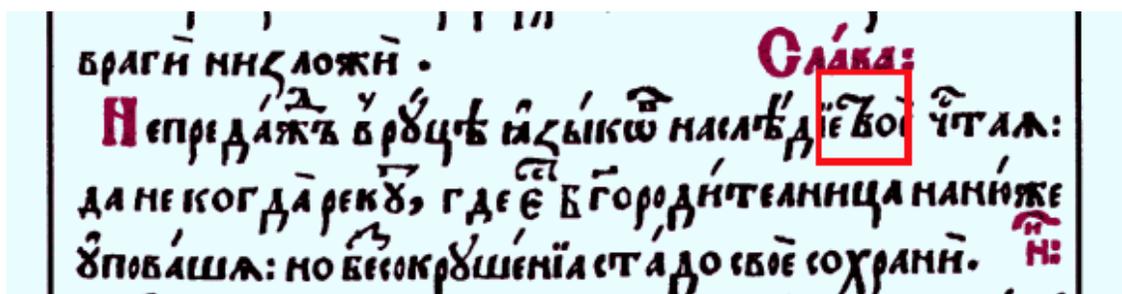


Figure 18: Cyrillic Small Letter Te (U+0442; boxed in black) and Cyrillic Small Letter Three-Legged Te (proposed at U+1C85; boxed in red). The Three-Legged form is the more prevalent form in this manuscript, and is used by default. Source: *Kanonnik* (Book of the Canons), a Poluustav manuscript written in 1616.

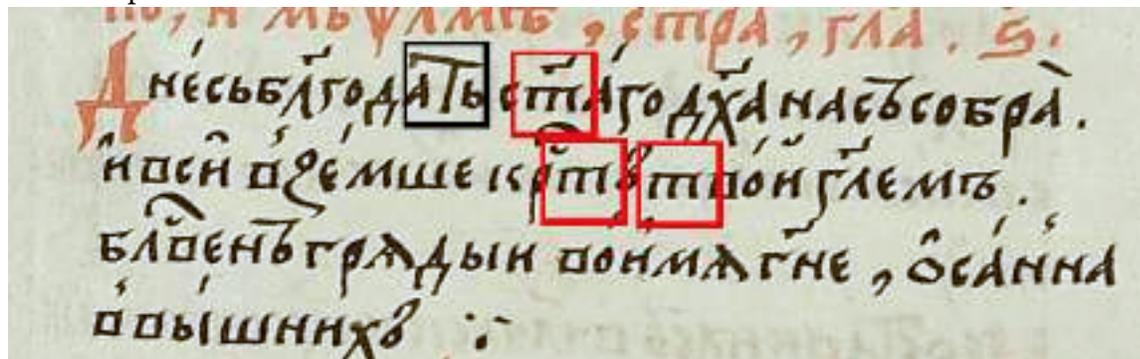


Figure 19: Cyrillic Small Letter Te (U+0442; boxed in black) and Cyrillic Small Letter Three-Legged Te (proposed at U+1C85; boxed in red). Observe that the Three-Legged form is used in initial position except in words of a Greek origin (e.g., *тропарѣ*, from the Greek *τροπάρια*). Source: *Flowery Triodion*, Lvov, 1642.

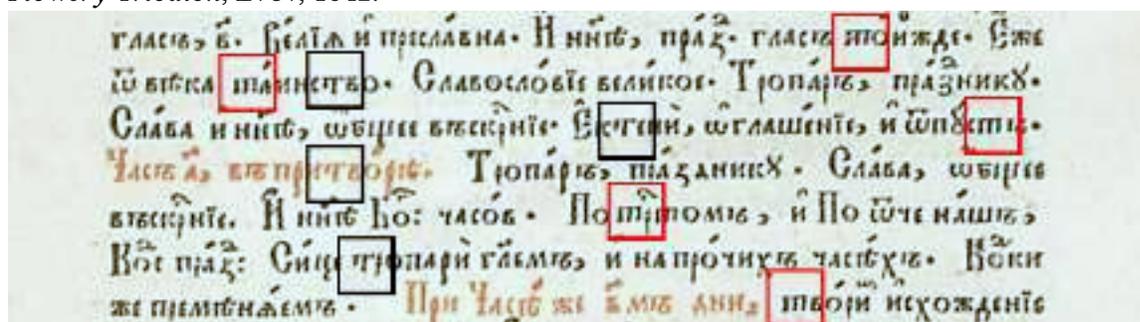


Figure 20: Cyrillic Small Letter Te (U+0442; boxed in black) and Cyrillic Small Letter Three-Legged Te (proposed at U+1C85; boxed in red). Observe that the Three-Legged form is used in medial position except in words of a Greek origin (e.g., *επιτιμίου*, from the Greek *ἐπιτίμια*). Source: *Homilies of St. John Chrysostom*, Polish-Lithuanian Commonwealth, c. 1600

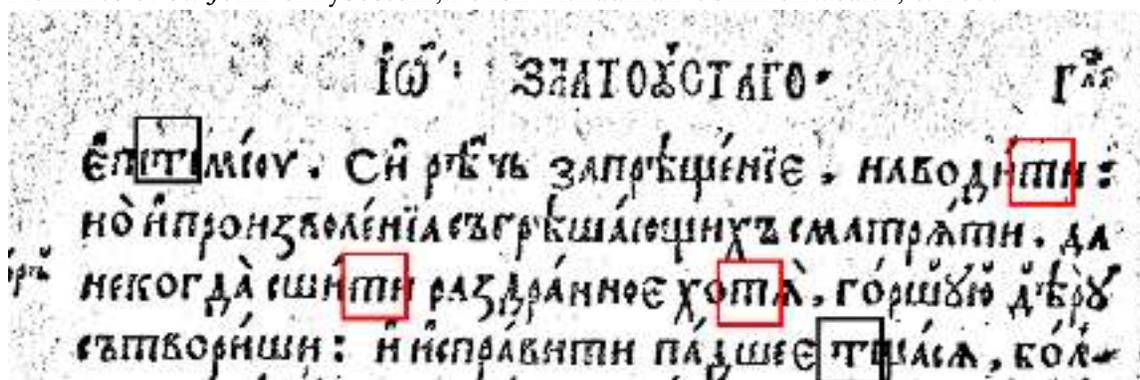


Figure 24: Cyrillic Small Letter Yat (U+0463; boxed in black) and Cyrillic Small Letter Tall Yat (proposed at U+1C87; boxed in red). Source: Gospel Book published by the Anonymous Press, Moscow, 1553.

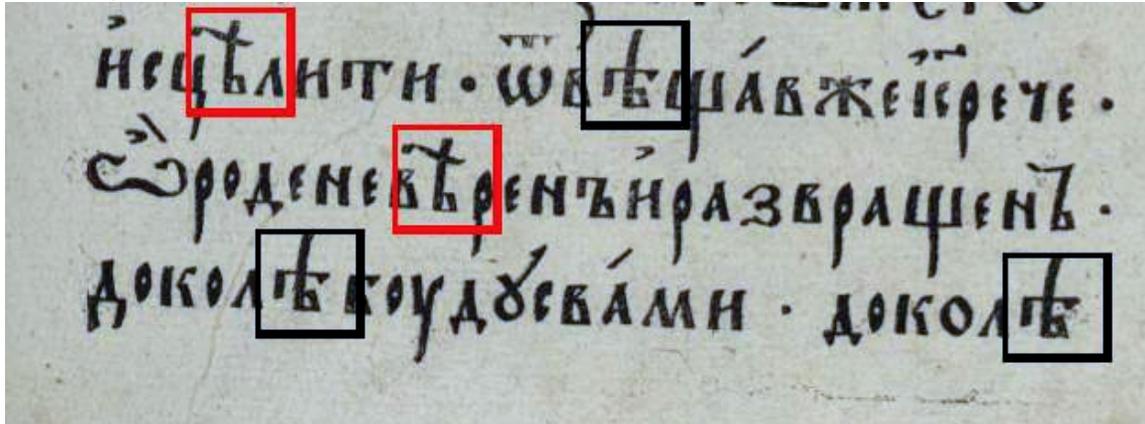
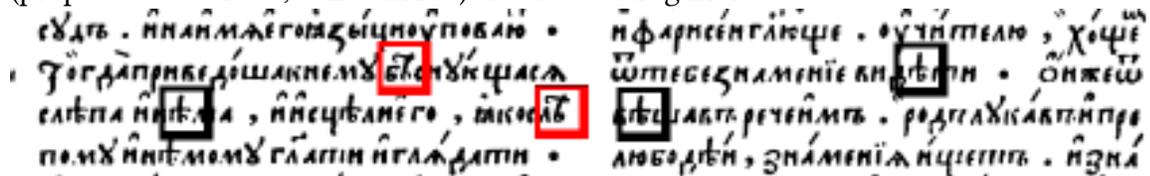


Figure 25: Cyrillic Small Letter Yat (U+0463; boxed in black) and Cyrillic Small Letter Tall Yat (proposed at U+1C87; boxed in red). Source: Ostrog Bible.



References

- Karsky, E. F. (1979). *Славянская Кирилловская Палеография*. Moscow: Nauka Press.
- Meyendorff, P. (1985). The liturgical reforms of Peter Moghila: A new look. *St Vladimir's Theological Quarterly* 29(2), 101–114.
- Nemirovsky, E. L. (2003). *История славянского кирилловского книгопечатания XV- начала XVII века*. Moscow, Russia: Nauka.
- Podokshin, S. A. (1981). *Франциск Скорина*. Moscow, Russia: Mysl'.
- Uspensky, B. A. (1987). *История Русского Литературного Языка (XI-XVII вв.)*. München: Verlag Otto Sagner.
- Uspensky, B. A. (1994). Раскол и культурный конфликт XVII века. In *Избранные труды. Том 1. Семиотика истории. Семиотика культуры*, pp. pp. 333–336. Moscow.
- Uspensky, N. D. (1975). Коллизия двух богословий в исправлении русских богослужебных книг в XVII веке. *Богословские труды* 13, pp. 148–171.

Appendix

Cyrillic Extended-C (Proposed)

8	
ѐ U+1C80	U+1C80: CYRILLIC SMALL LETTER ROUNDED VE
ё U+1C81	U+1C81: CYRILLIC SMALL LETTER LONG-LEGGED DE
ѐ U+1C82	U+1C82: CYRILLIC SMALL LETTER NARROW O
ѐ U+1C83	U+1C83: CYRILLIC SMALL LETTER WIDE ES
ѐ U+1C84	U+1C84: CYRILLIC SMALL LETTER TALL TE
ѐ U+1C85	U+1C85: CYRILLIC SMALL LETTER THREE-LEGGED TE
ѐ U+1C86	U+1C86: CYRILLIC SMALL LETTER TALL HARD SIGN
ѐ U+1C87	U+1C87: CYRILLIC SMALL LETTER TALL YAT
ѐ U+1C88	U+1C88: CYRILLIC SMALL LETTER UNBLENDED UK
ѐ U+1C89	U+1C89: <not assigned>
ѐ U+1C8A	U+1C8A: <not assigned>
ѐ U+1C8B	U+1C8B: <not assigned>
ѐ U+1C8C	U+1C8C: <not assigned>
ѐ U+1C8D	U+1C8D: <not assigned>
ѐ U+1C8E	U+1C8E: <not assigned>
ѐ U+1C8F	U+1C8F: <not assigned>

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹.**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title: **Proposal to Encode Additional Cyrillic Characters used in Early Church Slavonic Printed Books**

2. Requester's name: Aleksandr Andreev, Yuri Shardt and Nikita Simmons

3. Requester type (Member body/Liaison/Individual contribution): Individual contribution

4. Submission date: August 18, 2014

5. Requester's reference (if applicable): N/A

6. Choose one of the following:

This is a complete proposal:

(or) More information will be provided later:

B. Technical – General

1. Choose one of the following:

a. This proposal is for a new script (set of characters): YES

Proposed name of script: Cyrillic Extended-C

b. The proposal is for addition of character(s) to an existing block: NO

Name of the existing block: _____

2. Number of characters in proposal: 9

3. Proposed category (select one from below - see section 2.2 of P&P document):

A-Contemporary B.1-Specialized (small collection) B.2-Specialized (large collection)

C-Major extinct D-Attested extinct E-Minor extinct

F-Archaic Hieroglyphic or Ideographic G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided? YES

a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document? YES

b. Are the character shapes attached in a legible form suitable for review? YES

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? Aleksandr Andreev (aleksandr.andreev@gmail.com; http://www.ponomar.net/cy_support.html)

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): Hirmos Ponomar font licensed under GNU GPL by Aleksandr Andreev and Yuri Shardt

6. References:

a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? YES

b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? YES

7. Special encoding issues:

Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)? NO

8. Additional Information:

Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	YES
If YES explain	<i>Submitted as L2/13-153; revised based on comments from the UTC</i>
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	YES
If YES, with whom?	<i>Russian Old Rite communities, Slavonic typographers, academics</i>
If YES, available relevant documents:	<i>E-mail correspondence and person-to-person meetings</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	YES
Reference:	<i>See Section 3, Justification</i>
4. The context of use for the proposed characters (type of use; common or rare)	Rare
Reference:	<i>See Section 2, Proposed Characters</i>
5. Are the proposed characters in current use by the user community?	YES
If YES, where? Reference:	<i>In liturgical texts and academic publication. See Section 3.</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	YES
If YES, is a rationale provided?	YES
If YES, reference:	<i>See Section 4, Technical Issues</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	YES
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	<i>N/A</i>
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	NO
If YES, is a rationale for its inclusion provided?	
If YES, reference:	<i>N/A</i>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	YES
If YES, is a rationale for its inclusion provided?	YES
If YES, reference:	<i>See Section 2, Proposed Characters</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	NO
If YES, is a rationale for such use provided?	
If YES, reference:	<i>N/A</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	
If YES, reference:	<i>N/A</i>
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	NO
If YES, describe in detail (include attachment if necessary)	<i>N/A</i>
13. Does the proposal contain any Ideographic compatibility characters?	NO
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	<i>N/A</i>