

Proposal to Encode Combining Glagolitic Letters in Unicode

Aleksandr Andreev*, Heinz Miklas, Yuri Shardt

Section 1. Introduction

Glagolitic, also known as “Glagolitsa”, is an alphabetic writing system used to record Church Slavonic (ISO 639-2 code *cu*) and other Slavic languages. Originating in the 9th century, it is the earliest known Slavonic alphabet. The creation of the alphabet is attributed to the younger of the teachers of the Slavs, St. Cyril.

Glagolitic writing may be found in mediæval manuscripts and in printed liturgical books, mostly of a Croatian origin. In Bulgaria, Glagolitic was gradually replaced by the Cyrillic alphabet, and this Cyrillic alphabet was subsequently used also by other Slavs. For its part, the Glagolitic script has been preserved by some communities in Croatia even up to the present. Extant Glagolitic texts are of enormous value to linguists, palæographers, and scholars of liturgy.

Support for Glagolitic in the Unicode standard is required for two purposes. First, contemporary specialists need to be able to typographically represent mediæval texts written in the Glagolitic script, both in printed matter (such as academic publications) and in an electronic format (for use with computer analysis, such as string comparison, wordlist generation and searching). To this end, computer fonts that contain the repertoire of Glagolitic characters must be created. Second, owing to the close relationship between the Cyrillic and Glagolitic writing systems, scholars have traditionally represented Glagolitic texts also in Cyrillic transcription. To facilitate the transliteration process, an encoding model that parallels the model for the Cyrillic script needs to be available for Glagolitic.

The base repertoire of Glagolitic characters has been included in the Unicode standard since version 4.0. Nonetheless, this repertoire is incomplete because it lacks combining Glagolitic letters. Such combining letters exist in the Glagolitic script and play a function that is analogous to their role in Cyrillic – that is, they are used in abbreviations that are either space saving devices (for example, commonly written words are often abbreviated) or in *nomina sacra*. For full support of the Glagolitic writing system in Unicode, as well as for proper interoperability between the implementations of the Glagolitic and Cyrillic scripts, we propose the encoding of these combining characters in an additional block entitled *Glagolitic Extended-A*.

Section 2. Proposed Characters

The following table contains examples of combining Glagolitic letters that occur in various Glagolitic manuscripts and in printed literature. We propose to encode the characters as one block, in the same codepoint order as the base Glagolitic letters encoded at U+2C00 and following. This allows for simple computer manipulation of Glagolitic characters, as well as leaving some encoding positions empty to be used in the unlikely instance that additional combining characters are discovered by researchers and need to be encoded.

* Corresponding author, aleksandr.andreev@gmail.com.

Name	Codepoint	Appearance	Location in Sources
Combining Glagolitic Letter Azu	U+1E000	ⱦ	<i>Srez.</i> , p. 42; <i>MissSin</i> , 13v18, 44r16-17, 45v20; <i>PsDem</i> , 27v6, 42r7, 118r5
Combining Glagolitic Letter Buki	U+1E001	Ⱨ	<i>Srez.</i> , p. 254; <i>MissSin</i> , 38r13, 49v20
Combining Glagolitic Letter Vede	U+1E002	ⱨ	<i>Srez.</i> , p. 59; <i>MissSin</i> , 7r2, 22r10, 40v10, 52r17; <i>PsDem</i> , 104r8
Combining Glagolitic Letter Glagoli	U+1E003	Ⱪ	<i>Srez.</i> , p. 254; <i>PsDem</i> , 117r1
Combining Glagolitic Letter Dobro	U+1E004	ⱪ	<i>Srez.</i> , p. 228; <i>MissSin</i> , 54v14; <i>PsDem</i> , 21v1, 42r8, 78r19, 131r9
Combining Glagolitic Letter Yestu	U+1E005	ⱬ	<i>Srez.</i> , p. 59; <i>MissSin</i> , 29r20, 46r12
Combining Glagolitic Letter Zhivete	U+1E006	Ɑ	<i>Srez.</i> , p. 84
Combining Glagolitic Letter Zemlja	U+1E008	Ɐ	<i>EuchSinV</i> , 103r16m
Combining Glagolitic Letter Izhe	U+1E009	Ɒ	<i>Srez.</i> , p. 82
Combining Glagolitic Letter Initial Izhe	U+1E00A	ⱱ	<i>PsSinV</i> , 177r18
Combining Glagolitic Letter I	U+1E00B	Ⱳ	<i>Srezn.</i> , p. 82
Combining Glagolitic Letter Djervi	U+1E00C	ⱳ	<i>Srezn.</i> , p. 84; <i>MissSin</i> , 18v6
Combining Glagolitic Letter Kako	U+1E00D	ⱴ	<i>Srezn.</i> , p. 224; <i>PsDem</i> , 126r2
Combining Glagolitic Letter Ljudie	U+1E00E	Ⱶ	<i>Srezn.</i> , p. 40, p. 42; <i>PsDem</i> , 5v1; <i>MissSin.</i> , 30v22, 51r15
Combining Glagolitic Letter Myslite	U+1E00F	ⱶ	<i>Srezn.</i> , p. 224; <i>PsDem</i> , 50v15, 105r19, 113v3; <i>MissSin</i> , 13r18
Combining Glagolitic Letter Nashi	U+1E010	ⱷ	<i>Srezn.</i> , p. 59; <i>PsDem</i> , 21v3; <i>MissSin</i> , 20v23
Combining Glagolitic Letter Onu	U+1E011	ⱸ	<i>PsDem</i> , 126r2.; <i>MissSin</i> , 43v14, 43v21-2
Combining Glagolitic Letter Pokoji	U+1E012	ⱹ	<i>Srezn.</i> , p. 42, p. 248; <i>MissSin</i> , 13r18, 13v10
Combining Glagolitic Letter Ritsi	U+1E013	ⱺ	<i>Srezn.</i> , p. 42; <i>MissSin</i> , 35v15
Combining Glagolitic Letter Slovo	U+1E014	ⱻ	<i>Srezn.</i> , p. 228; <i>PsDem</i> , 126r24, 128v7, 126r24, 126v18
Combining Glagolitic Letter Tvrido	U+1E015	ⱼ	<i>Srezn.</i> , p. 36, p. 42, 59; <i>PsDem</i> , 118r4; <i>MissSin</i> , 18v3, 17v12, 45v20, 46r19

Name	Codepoint	Appearance	Location in Sources
Combining Glagolitic Letter Uku	U+1E016	ⱸ	<i>MissSin</i> , 43v14
Combining Glagolitic Letter Fritu	U+1E017	ⱹ	<i>MissSin</i> , 25r13, 29r14, 33r23, 40r15, 22r10
Combining Glagolitic Letter Heru	U+1E018	ⱺ	<i>MissSin</i> , 19r12, 13v15, 17v14
Combining Glagolitic Letter Shta	U+1E01B	ⱻ	<i>MissSin</i> , 58r6, 72r/v8
Combining Glagolitic Letter Tsi	U+1E01C	ⱼ	Srezn., p. 59; <i>MissSin</i> , 19r12
Combining Glagolitic Letter Chrivi	U+1E01D	ⱽ	<i>MissSin</i> , 53r5, 45r9-10, 39v13
Combining Glagolitic Letter Sha	U+1E01E	Ȿ	Srezn., p. 224; <i>MissSin</i> , 21r12, 13r1, 17r18, 22v18, 24r23, 24(15)v8
Combining Glagolitic Letter Yeru	U+1E01F	Ɀ	<i>MissSin</i> , 18v6
Combining Glagolitic Letter Yeri	U+1E020	Ⳁ	Srezn., p. 42
Combining Glagolitic Letter Yati	U+1E021	ⳁ	Srezn., p. 248; <i>PsDem</i> , 83r16; <i>MissSin</i> , 40r15
Combining Glagolitic Letter Yu	U+1E023	ⳃ	Srezn., p. 254
Combining Glagolitic Letter Small Yus	U+1E024	Ⳅ	<i>EuchSinV</i> , 32v17, 51r11
Combining Glagolitic Letter Yo	U+1E026	Ⳇ	Does not exist as a single character, but is a component of U+E029.
Combining Glagolitic Letter Iotated Small Yus	U+1E027	Ⳅ̇	Srezn., p. 248; <i>MissSin</i> , 19r12
Combining Glagolitic Letter Big Yus	U+1E028	Ⳇ̇	Mansvetov, p. 362 (given by Mansvetov in Cyrillic transcription only)
Combining Glagolitic Letter Iotated Big Yus	U+1E029	Ⳇ̇̇	<i>EuchSinV</i> , 5r10, 17r1, 17v18, 28v3, 29r18 etc.
Combining Glagolitic Letter Fita	U+1E02A	ⳇ	<i>Assem</i> , 125v5, 149v25; (<i>EuchSinV</i> , 83v14, 84v15, 85v4 for /f/)

The following entries are proposed for addition to UnicodeData.txt:

```

1E000;COMBINING GLAGOLITIC LETTER AZU;Mn;230;NSM;;;;;N;;;;;
1E001;COMBINING GLAGOLITIC LETTER BUKI;Mn;230;NSM;;;;;N;;;;;
1E002;COMBINING GLAGOLITIC LETTER VEDE;Mn;230;NSM;;;;;N;;;;;
1E003;COMBINING GLAGOLITIC LETTER GLAGOLI;Mn;230;NSM;;;;;N;;;;;
1E004;COMBINING GLAGOLITIC LETTER DOBRO;Mn;230;NSM;;;;;N;;;;;

```


Section 4. Justification for Encoding

In this section, we explain the rationale behind encoding combining Glagolitic letters in the Unicode standard. In summary, we argue that because combining Cyrillic characters have already been encoded in Unicode, correct interoperability between the two scripts demands that combining Glagolitic characters be also encoded.

4.1 Distinction and Use of Combining Characters in Church Slavonic

As we stated in the Introduction section above, superscription in Church Slavonic is used in two instances: in abbreviations (for example, the word *милостѣ* (mercy) is often written as *млѣтѣ*) and in *nomina sacra* (for example, the spelling *гѣ* (Lord) is used when it refers to God and the spelling *госпѣ* is used when it refers to a secular ruler (a lord), much the same way that capitalization is used in many modern languages).¹ Thus, superscription is a required feature of writing Church Slavonic. Unlike in modern English and other languages where superscription is a stylistic embellishment (e.g., in writing “2nd” as opposed to “2nd”), the superscripted characters in Church Slavonic are combining characters that act like true diacritical marks. In particular, these characters (both in the Cyrillic and Glagolitic scripts) are non-spacing characters, while the “n” and “d” in writing “2nd” in English are spacing characters. Handling such non-spacing characters by positioning spacing characters over a base character would not be correct from the standpoint of text processing in the Unicode standard. In fact, as far as text processing is concerned, the combining characters in Church Slavonic are in no way different from any of the other diacritical marks already encoded.

In his review of this proposal ([L2/14-103](#)), David Birnbaum concedes that “standard modern ChSl [Church Slavonic] orthography does require superscript letters in some words.” But then he goes on to write, “I would have regarded the use of the “wrong” letter as culturally incorrect but nonetheless informationally adequate.” We strongly disagree with this premise. In many instances, using the inline letter instead of the combining letter is not only “culturally incorrect” but also “informationally inadequate”. For example, the sequence *ѡг[ъ]* means “god” (a pagan deity) while the sequence *ѡѣ* (with the combining letter Ge) is an abbreviation for *ѡгоспѣдиченъ* (“theotokion” – a type of liturgical hymn). While one could write *ѡѣ* by using markup-level superscription, this “r” is still a spacing character. One could use kerning at the font level to force the “r” to position over the “o”, but such an approach is not correct from the standpoint of text processing. Moreover, under such an approach, the correct meaning of a text stream would not only be determined by its characters but also by markup (or formatting) and by font-level attributes, which cannot be exchanged between users in a plain text context.

In addition, it is also incorrect to write all Church Slavonic words in their full, unabbreviated form (resolving all abbreviations). As we have seen above, *гѣ* and *госпѣ* have two different meanings; thus, writing *гѣ* as *госпѣ* is “informationally inadequate,” not just “culturally incorrect”.

It is true that strict orthographic conventions did not take shape in Church Slavonic until after 1700 (with the publication of the Elizabeth Bible in 1751) and that Church Slavonic writing of earlier recensions (especially before the advent of the printing press) demonstrates a greater degree of

¹ Throughout we present examples in the Cyrillic script, since it is more familiar to the reader. We will then demonstrate that all of the arguments also hold true for the Glagolitic script because of the relationship that exists between the two scripts.

leeway in spelling. Nonetheless, wherever combining letters occur in Church Slavonic, they are always treated as non-spacing marks. Treating them differently in computer-encoded text would be an unreasonable limitation.

In fact, combining characters are already encoded in Unicode for a variety of writing systems. In addition to the combining characters used for Cyrillic, Unicode includes a variety of combining characters used in writing classical Arabic (for example, Honorifics and Koranic annotation signs; for a discussion of these, see [L2/01-425](#)); and the various combining Latin characters used for the representation of mediæval texts (see [ISO/IEC JTC 1/SC 2/WG 2 N2266](#) and [L2/06-027](#)). In their proposal for combining Latin characters, Everson *et. al.* argue that these characters are necessary to make possible a representation of mediæval text that “does not entail the replacement or the distortion of the original character set.” Similarly, a distortion of the Glagolitic character set is undesirable and so combining Glagolitic characters are needed.

4.2 Interoperability of Cyrillic and Glagolitic

The Cyrillic and Glagolitic scripts are closely related. Combining characters have already been encoded for Cyrillic and a methodology needs to exist for the simple automated conversion of texts between Cyrillic and Glagolitic. Thus, combining Glagolitic letters should also be encoded.

As we have pointed out in the Introduction section, scholars believe that Glagolitic was the first script used to record Church Slavonic. After the introduction of the Cyrillic script, it gradually replaced Glagolitic in most Slavic cultures. Nonetheless, Cyrillic and Glagolitic scripts are often used interchangeably in the academic community, and it is quite common to publish Glagolitic textual sources in Cyrillic transcription (see Figure 14). Because of this interoperability of Cyrillic and Glagolitic, computer software needs to be able to unambiguously convert (transliterate) between the two writing systems. In the Table below, we present the transliteration scheme commonly used and due to Jagic (1879):

Glagolitic			Cyrillic		
Name	Codept.	Disp.	Name	Codept.	Disp.
Glagolitic Letter Azu	U+2C30	ⱥ	Cyrillic Letter A	U+0430	А
Glagolitic Letter Buki	U+2C31	ⱦ	Cyrillic Letter Be	U+0431	Б
Glagolitic Letter Vede	U+2C32	Ⱨ	Cyrillic Letter Ve	U+0432	В
Glagolitic Letter Glagoli	U+2C33	ⱨ	Cyrillic Letter Ge	U+0433	Г
Glagolitic Letter Dobro	U+2C34	Ⱪ	Cyrillic Letter De	U+0434	Д
Glagolitic Letter Yestu	U+2C35	ⱪ	Cyrillic Letter Ie	U+0435	Е
Glagolitic Letter Zhivete	U+2C36	ⱬ	Cyrillic Letter Zhe	U+0436	Ж
Glagolitic Letter Dzelo	U+2C37	Ɑ	Cyrillic Letter Dze	U+0455	З
Glagolitic Letter Zemlja	U+2C38	Ɱ	Cyrillic Letter Ze	U+0437	З
Glagolitic Letter Izhe	U+2C39	Ɐ	Cyrillic Letter Ukrainian I	U+0456	І
Glagolitic Letter Initial Izhe	U+2C3A	Ɒ	Cyrillic Letter Iota	U+A647	Ѳ
Glagolitic Letter I	U+2C3B	ⱱ	Cyrillic Letter I	U+0438	И

Glagolitic			Cyrillic		
Glagolitic Letter Djervi	U+2C3C	Ѣ	Cyrillic Letter Djerv	U+A649	Ѣ
Glagolitic Letter Kako	U+2C3D	Ѧ	Cyrillic Letter Ka	U+043A	Ѧ
Glagolitic Letter Ljudie	U+2C3E	Ѧ	Cyrillic Letter El	U+043B	Ѧ
Glagolitic Letter Myslite	U+2C3F	Ѧ	Cyrillic Letter Em	U+043C	Ѧ
Glagolitic Letter Nashi	U+2C40	Ѧ	Cyrillic Letter En	U+043D	Ѧ
Glagolitic Letter Onu	U+2C41	Ѧ	Cyrillic Letter O	U+043E	Ѧ
Glagolitic Letter Pokoji	U+2C42	Ѧ	Cyrillic Letter Pe	U+043F	Ѧ
Glagolitic Letter Ritsi	U+2C43	Ѧ	Cyrillic Letter Er	U+0440	Ѧ
Glagolitic Letter Slovo	U+2C44	Ѧ	Cyrillic Letter Es	U+0441	Ѧ
Glagolitic Letter Tvrido	U+2C45	Ѧ	Cyrillic Letter Te	U+0442	Ѧ
Glagolitic Letter Uku	U+2C46	Ѧ	Cyrillic Letter O Cyrillic Letter U	U+043E U+0443	Ѧ
Glagolitic Letter Fritu	U+2C47	Ѧ	Cyrillic Letter Ef	U+0444	Ѧ
Glagolitic Letter Heru	U+2C48	Ѧ	Cyrillic Letter Ha	U+0445	Ѧ
Glagolitic Letter Otu	U+2C49	Ѧ	Cyrillic Letter Omega	U+0461	Ѧ
Glagolitic Letter Shta	U+2C4B	Ѧ	Cyrillic Letter Shcha	U+0449	Ѧ
Glagolitic Letter Tsi	U+2C4C	Ѧ	Cyrillic Letter Tse	U+0446	Ѧ
Glagolitic Letter Chrivi	U+2C4D	Ѧ	Cyrillic Letter Che	U+0447	Ѧ
Glagolitic Letter Sha	U+2C4E	Ѧ	Cyrillic Letter Sha	U+0448	Ѧ
Glagolitic Letter Yeru	U+2C4F	Ѧ	Cyrillic Letter Hard Sign	U+044A	Ѧ
Glagolitic Letter Yeru, Glagolitic Letter Izhe	U+2C4F U+2C39	Ѧ	Cyrillic Letter Yeru with Back Yer	U+A651	Ѧ
Glagolitic Letter Yeri	U+2C50	Ѧ	Cyrillic Letter Soft Sign	U+044C	Ѧ
Glagolitic Letter Yati	U+2C51	Ѧ	Cyrillic Letter Yat	U+0463	Ѧ
Glagolitic Letter Yu	U+2C53	Ѧ	Cyrillic Letter Yu	U+044E	Ѧ
Glagolitic Letter Small Yus	U+2C54	Ѧ	Cyrillic Letter Little Yus	U+0467	Ѧ
Glagolitic Letter Iotated Small Yus	U+2C57	Ѧ	Cyrillic Letter Iotified Little Yus	U+0469	Ѧ
Glagolitic Letter Big Yus	U+2C58	Ѧ	Cyrillic Letter Big Yus	U+046B	Ѧ
Glagolitic Letter Iotated Big Yus	U+2C59	Ѧ	Cyrillic Letter Iotified Big Yus	U+046D	Ѧ
Glagolitic Letter Fita	U+2C5A	Ѧ	Cyrillic Letter Fita	U+0473	Ѧ
Glagolitic Letter Izhitsa	U+2C5B	Ѧ	Cyrillic Letter Izhitsa	U+0475	Ѧ

The Table presents the standard transliteration scheme for the main letters of the Cyrillic and Glagolitic scripts. However, we contend that an automated conversion (transliteration) algorithm also needs to correctly handle diacritical marks and combining letters present in the two scripts.

In the encoding model for the Cyrillic script, the combining Cyrillic letters are already available. In their responses to this proposal ([L2/14-103](#) and [L2/14-165](#)), Cleminson and Birnbaum present an overall negative view of the use of combining characters for Cyrillic. Cleminson believes that only

those combining Cyrillic letters that occur in modern Church Slavonic should have been encoded and Birnbaum writes that he “cannot now endorse the inclusion of any additional superscript Cyrillic or Glagolitic characters.”

In hindsight, we agree that the use of combining characters for Cyrillic superscription was not the best approach (although our criticism of this approach is based on different reasons than those presented by Cleminson and Birnbaum). In our view, this approach is too limiting and does not allow for the unambiguous representation of various complexities occurring in mediæval texts, such as, for example, the use of superscription over multiple base letters or the use of multiple combining letters over one base letter. Also, we regret that the standardization of Glagolitic and Cyrillic was not handled simultaneously, leading to discrepancies in the encoding schemes. Nonetheless, given Unicode's stability policy, there is no use now to criticize an existing implementation. Rather, we desire to make the existing implementation more useful. Thus, because combining Cyrillic letters are encoded, we propose that combining Glagolitic letters also need to be encoded so that a meaningful transliteration scheme can be designed. An approach where Cyrillic superscription is handled via combining characters while Glagolitic superscription is rendered using markup or stylistic alternatives would not make such a transliteration scheme possible. But, as we have pointed out, since it is common to present Glagolitic texts in Cyrillic transcription, the availability of such a scheme is necessary. The Table below presents the proposed transliteration scheme for combining letters. Note, again, that all codepoints for Glagolitic characters are provisional.

Glagolitic			Cyrillic		
Name	Codept.	Disp.	Name	Codept.	Disp.
Glagolitic Combining Letter Azu	U+1E000	ⱦ	Cyrillic Combining Letter A	U+2DF6	ⱦ
Glagolitic Combining Letter Buki	U+1E001	Ⱨ	Cyrillic Combining Letter Be	U+2DE0	Ⱨ
Glagolitic Combining Letter Vede	U+1E002	ⱨ	Cyrillic Combining Letter Ve	U+2DE1	ⱨ
Glagolitic Combining Letter Glagoli	U+1E003	Ⱪ	Cyrillic Combining Letter Ge	U+2DE2	Ⱪ
Glagolitic Combining Letter Dobro	U+1E004	ⱪ	Cyrillic Combining Letter De	U+2DE3	ⱪ
Glagolitic Combining Letter Yestu	U+1E005	ⱬ	Cyrillic Combining Letter Ie	U+2DF7	ⱬ
Glagolitic Combining Letter Zhivete	U+1E006	Ɑ	Cyrillic Combining Letter Zhe	U+2DE4	Ɑ
Glagolitic Combining Letter Zemlja	U+1E008	Ɐ	Cyrillic Combining Letter Ze	U+2DE5	Ɐ
Glagolitic Combining Letter Izhe	U+1E009	Ɒ	Cyrillic Combining Letter Yi	U+A676	Ɒ
Glagolitic Combining Letter Initial Izhe	U+1E00A	ⱱ	Cyrillic Combining Letter Iota	N/A	
Glagolitic Combining Letter I	U+1E00B	Ⱳ	Cyrillic Combining Letter I	U+A675	Ⱳ
Glagolitic Combining Letter Djervi	U+1E00C	ⱳ	Cyrillic Combining Letter Djerv	U+2DF8	ⱳ
Glagolitic Combining Letter Kako	U+1E00D	ⱴ	Cyrillic Combining Letter Ka	U+2DE6	ⱴ
Glagolitic Combining Letter	U+1E00E	Ⱶ	Cyrillic Combining Letter El	U+2DE7	Ⱶ

Glagolitic			Cyrillic		
Ljudie					
Glagolitic Combining Letter Myslite	U+1E00F	Ɱ	Cyrillic Combining Letter Em	U+2DE8	Ѡ
Glagolitic Combining Letter Nashi	U+1E010	Ɐ	Cyrillic Combining Letter En	U+2DE9	ѡ
Glagolitic Combining Letter Onu	U+1E011	Ɒ	Cyrillic Combining Letter O	U+2DEA	Ѣ
Glagolitic Combining Letter Pokoji	U+1E012	ⱱ	Cyrillic Combining Letter Pe	U+2DEB	ѣ
Glagolitic Combining Letter Ritsi	U+1E013	Ⱳ	Cyrillic Combining Letter Er	U+2DEC	Ѥ
Glagolitic Combining Letter Slovo	U+1E014	ⱳ	Cyrillic Combining Letter Es	U+2DED	ѥ
Glagolitic Combining Letter Tvrido	U+1E015	ⱴ	Cyrillic Combining Letter Te	U+2DEE	Ѧ
Glagolitic Combining Letter Uku	U+1E016	Ⱶ	Cyrillic Combining Letter Monograph Uk	U+2DF9	ѧ
Glagolitic Combining Letter Fritu	U+1E017	ⱶ	(Cyrillic Combining Letter Ef)	(U+A69E)	Ѩ
Glagolitic Combining Letter Heru	U+1E018	ⱷ	Cyrillic Combining Letter Ha	U+2DEF	ѩ
Glagolitic Combining Letter Shta	U+1E01B	ⱸ	Cyrillic Combining Letter Shcha	U+2DF3	Ѭ
Glagolitic Combining Letter Tsi	U+1E01C	ⱹ	Cyrillic Combining Letter Tse	U+2DF0	ѭ
Glagolitic Combining Letter Chrivi	U+1E01D	ⱺ	Cyrillic Combining Letter Che	U+2DF1	Ѯ
Glagolitic Combining Letter Sha	U+1E01E	ⱻ	Cyrillic Combining Letter Sha	U+2DF2	ѯ
Glagolitic Combining Letter Yeru	U+1E01F	ⱼ	Cyrillic Combining Letter Hard Sign	U+A678	Ѱ
Glagolitic Combining Letter Yeri	U+1E020	ⱽ	Cyrillic Combining Letter Soft Sign	U+A67A	Ѳ
Glagolitic Combining Letter Yati	U+1E021	Ȿ	Cyrillic Combining Letter Yat	U+2DFA	ѳ
Glagolitic Combining Letter Yu	U+1E023	Ɀ	Cyrillic Combining Letter Yu	U+2DFB	Ѵ
Glagolitic Combining Letter Small Yus	U+1E024	Ⳁ	Cyrillic Combining Letter Little Yus	U+2DFD	Ѷ
Glagolitic Combining Letter Iotated Small Yus	U+1E027	ⳁ	Cyrillic Combining Letter Iotified Little Yus	N/A ²	
Glagolitic Combining Letter Big Yus	U+1E028	Ⳃ	Cyrillic Combining Letter Big Yus	U+2DFE	Ѹ
Glagolitic Combining Letter Iotated Big Yus	U+1E029	ⳃ	Cyrillic Combining Letter Iotified Big Yus	U+2DFF	ѹ
Glagolitic Combining Letter Fita	U+1E02A	Ⳅ	Cyrillic Combining Letter Fita	U+2DF4	Ѻ

4.4 Collation

One should note that under this encoding methodology, for example, the strings **НАШНХ** and **НАШ^X**

² This character will be proposed for encoding by the authors in a separate document.

are not equivalent at the codepoint level even though under certain circumstances one would like to treat them as the same string. However, the proper way to compare these strings is not via direct (codepoint by codepoint) comparison but rather via the use of the Unicode Collation Algorithm (UCA), which allows for multi-level comparison codes. Under an appropriate UCA collation table, the characters X and X^{X} could be given the same primary weights but different secondary weights (which is in fact what is implemented for Cyrillic and what we propose for Glagolitic). In a context where the above strings need to be treated as identical one can simply perform a comparison ignoring the secondary weights, while in the context where these strings need to be different, one can compare them at the secondary level. In any case, comparing any two UTF strings codepoint by codepoint is not advisable from the standpoint of text processing.

4.5 Urgency

A number of projects are currently active in the digital storage and presentation of Glagolitic texts. The *Manuscript: Ancient Slavonic Literary Monuments* project affiliated with Udmurt State University and Izhevsk State Technical University (<http://www.manuscripts.ru/>) has made a number of Glagolitic manuscripts available in electronic format, including the Kiev Missal and Codex Assemanius. These manuscripts are presented on the project webpage both in the Glagolitic original and in Cyrillic transcription. Because combining Glagolitic characters are not available in the Unicode standard, the authors encode the necessary characters in the PUA. The encoding of these characters – as well as a standardized conversion methodology between Glagolitic and Cyrillic – would allow the data provided by this project to be Unicode compatible.

Another project dedicated to the study of Glagolitic texts is *The Enigma of the Sinaitic Glagolitic Tradition Project* (<http://www.caa.tuwien.ac.at/cvl/research/sinai/>), an interdisciplinary collaboration of the Institute of Slavic Studies, University of Vienna, and other scholars. The project is devoted to the study of Slavonic Glagolitic manuscripts discovered at St. Catherine's Monastery at the foot of Mt. Sinai, Egypt, and the development of technical means for the preservation and study of such texts. The project uses computer methods, including spectral imaging and optical character recognition (OCR), to study these manuscripts with a goal to publish both critical and facsimile editions. The first volume of the *Glagolitica Sinaitica* series, a facsimile of the Psalter of Demetrius (*Psalterium Demetrii Sinaitici*), was published in 2012. The standardization of Glagolitic characters and automated Glagolitic to Cyrillic conversion is urgently needed by this project for continued work on critical editions, commentaries, and indexes.

Section 5. Examples


- 
 1. 300-655P330-65 300-655P33: 30+
 2. 3PΔ336 P30Δ 333-655P6 00P35 P6Δ

Figure 1: Combining Glagolitic Letter Azu (boxed in red). Source: Srezn.

06. (верхнюю половину страницы занимает большая
 записка.)

 1 333 300+ 00+ 33 P6Δ0333P35

 бббб зз зз зз зз зз

Figure 2: Combining Glagolitic Letters Buky (boxed in red); Glagoli (boxed in black); Ljudije (boxed in blue); Tvrido (boxed in yellow); and Yu (boxed in green). Note the use of the Pokrytie. Source: Srezn.


15. 33333333 333 33 3333

 ΔΔΔ ΔΔ ΔΔΔΔΔΔΔΔ ΔΔΔΔ
 ΔΔΔΔΔΔΔΔΔΔ ΔΔΔΔ 33

Figure 3: Combining Glagolitic Letters Dobro (boxed in blue); Slovo (boxed in red) and Tvrido (boxed in green). Source: Srezn.

33333333333. 3 P33+P33Δ

 3 ΔP333333 333 3333
 ΔΔΔΔ ΔΔΔ P33333333

Figure 4: Combining Glagolitic Letter Kako (boxed in red). Note the use of the Pokrytie. Source: Srezn.


6.

 P3333 333 33333. P3333 333+ 00333Δ. P33P3333
 P3333333 00333 33333. P 3333333 3333Δ 3333. P
 P33P33333 3333333 0033333. P 3333333P 333 33333+ 3333Δ.

Figure 5: Combining Glagolitic Letter Heru (boxed in orange). Source: Srezn.

References

- Assem*: Codex Assemanianus (Vat. slav. 3 glag.), Macedonia, early 11th c. Available online (printed in *Asemanievo evangelie. Faksimilno izdanie*. Sofia: Nauka i izkustvo, 1981).
- EuchSinV*: Euchologium Sinaiticum, old part (Sin. slav. 37 & SPb. RNB, Glag. 3), Bulgaria, 11th c.; facsimile edition: NAHTIGAL, R. *Euchologium Sinaiticum. Starocerkvenoslovanski glagolski spomenik. I. Fotografski posnetek*. Ljubljana: AZU v Ljubljani, Učiteljska tiskarna, 1941.
- Everson: EVERSON, Michael *et. al.* "Proposal to add medievalist characters to the UCS". Working Group Document L2/06-027. 2006.
- Jagic: JAGIĆ, Vatroslav, *Quattuor Evangeliorum Codex Glagoliticus, olim Zographensis, nunc Petropolitanus*. Berlin: Apud Weidmannos, 1879.
- Mansvetov: MANSVETOV, I. D., "Evchologium Glagolski, Spomenik monastira Sinai brda. Izdao D-r Lavoslav Geitler. U Zagrebu, 1882. (Древнейший Славянский требник, изданный г. Гейтлером по рукописи, находящейся в Синайской библиотеке)". In *Прибавления к Творениям св. отцов*. 1883. vol. 32, part 1, pp. 347-390.
- MissSin*: Sinai Missal (Sin. slav. 5/N), Bulgaria, 11th c.; cf. MIKLAS, H. (ed.): *Glagolitica – Zum Ursprung der slavischen Schriftkultur* (ÖAW, Phil.-hist. Kl., Schriften der Balkan-Kommission, Philologische Abt. 41). Vienna: ÖAW, 2000, pp. 117-129.
- PsDem*: Psalter of Demetrius (Sin. slav. 3/N), Sinai, 11th c. Facsimile edition: *Psalterium Demetrii Sinaitici (monasterii sanctae Catharinae codex slav. 3/N), adiectis foliis medicinalibus*. Ad editionem phototypicam praeparaverunt M. Gau, D. Hürner, F. Kleber, M. Lettner, H. Miklas sub redactione Henrici Miklas. Cum praefationibus sacri monasterii atque Ioannis Tarnanidae (Glagolitica Sinaitica 1). Vienna: Holzhausen, 2012.
- PsSinV*: Psalterium Sinaiticum, old part (Sin. slav. 38), Sinai, 11th c., Facsimile edition: ALTBAUER, M.: *Psalterium Sinaiticum. An 11th Century Glagolitic Manuscript from St. Catherine's Monastery, Mt. Sinai*. Skopje: The Macedonian Academy of Sciences and Arts, Goce Delčev Publishing House, 1971.
- Srez.: SREZNEVSKIJ, I. I., *Древніе глаголическіе памятники, сравнительно съ памятниками кириллицы*, St. Petersburg: Типографія Императорской академіи наук, 1866.

Glagolitic Extended-A

0	1	2
Ɱ U+1E000	Ɱ U+1E010	Ɱ U+1E020
Ɱ U+1E001	Ɱ U+1E011	Ɱ U+1E021
Ɱ U+1E002	Ɱ U+1E012	Ɱ U+1E022
Ɱ U+1E003	Ɱ U+1E013	Ɱ U+1E023
Ɱ U+1E004	Ɱ U+1E014	Ɱ U+1E024
Ɱ U+1E005	Ɱ U+1E015	Ɱ U+1E025
Ɱ U+1E006	Ɱ U+1E016	Ɱ U+1E026
Ɱ U+1E007	Ɱ U+1E017	Ɱ U+1E027
Ɱ U+1E008	Ɱ U+1E018	Ɱ U+1E028
Ɱ U+1E009	Ɱ U+1E019	Ɱ U+1E029
Ɱ U+1E00A	Ɱ U+1E01A	Ɱ U+1E02A
Ɱ U+1E00B	Ɱ U+1E01B	Ɱ U+1E02B
Ɱ U+1E00C	Ɱ U+1E01C	Ɱ U+1E02C
Ɱ U+1E00D	Ɱ U+1E01D	Ɱ U+1E02D
Ɱ U+1E00E	Ɱ U+1E01E	Ɱ U+1E02E
Ɱ U+1E00F	Ɱ U+1E01F	Ɱ U+1E02F

U+1E000: Combining Glagolitic Letter Azu
 U+1E001: Combining Glagolitic Letter Buki
 U+1E002: Combining Glagolitic Letter Vede
 U+1E003: Combining Glagolitic Letter Glagoli
 U+1E004: Combining Glagolitic Letter Dobro
 U+1E005: Combining Glagolitic Letter Yestu
 U+1E006: Combining Glagolitic Letter Zhivete
 U+1E007: <not assigned>
 U+1E008: Combining Glagolitic Letter Zemlja
 U+1E009: Combining Glagolitic Letter Izhe
 U+1E00A: Combining Glagolitic Letter Initial Izhe
 U+1E00B: Combining Glagolitic Letter I
 U+1E00C: Combining Glagolitic Letter Djervi
 U+1E00D: Combining Glagolitic Letter Kako
 U+1E00E: Combining Glagolitic Letter Ljudie
 U+1E00F: Combining Glagolitic Letter Myslite

 U+1E010: Combining Glagolitic Letter Nashi
 U+1E011: Combining Glagolitic Letter Onu
 U+1E012: Combining Glagolitic Letter Pokoji
 U+1E013: Combining Glagolitic Letter Ritsi
 U+1E014: Combining Glagolitic Letter Slovo
 U+1E015: Combining Glagolitic Letter Tvrido
 U+1E016: Combining Glagolitic Letter Uku
 U+1E017: Combining Glagolitic Letter Fritu
 U+1E018: Combining Glagolitic Letter Heru
 U+1E019: <not assigned>
 U+1E01A: <not assigned>
 U+1E01B: Combining Glagolitic Letter Shta
 U+1E01C: Combining Glagolitic Letter Tsi
 U+1E01D: Combining Glagolitic Letter Chrivi
 U+1E01E: Combining Glagolitic Letter Sha
 U+1E01F: Combining Glagolitic Letter Yeru

 U+1E020: Combining Glagolitic Letter Yeri
 U+1E021: Combining Glagolitic Letter Yati
 U+1E022: <not assigned>
 U+1E023: Combining Glagolitic Letter Yu
 U+1E024: Combining Glagolitic Letter Small Yus
 U+1E025: <not assigned>
 U+1E026: Combining Glagolitic Letter Yo
 U+1E027: Combining Glagolitic Letter Iotated Small Yus
 U+1E028: Combining Glagolitic Letter Big Yus
 U+1E029: Combining Glagolitic Letter Iotated Big Yus
 U+1E02A: Combining Glagolitic Letter Fita
 U+1E02B: <not assigned>
 U+1E02C: <not assigned>
 U+1E02D: <not assigned>
 U+1E02E: <not assigned>
 U+1E02F: <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 Combining Glagolitic Letters in Unicode	
2. Requester's name:	<i>Aleksandr Andreev, Heinz Miklas and Yuri Shardt</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>July 8, 2014</i>	
5. Requester's reference (if applicable):	<i>N/A</i>	
6. Choose one of the following:		
This is a complete proposal:	<input checked="" type="checkbox"/>	
(or) More information will be provided later:	<input type="checkbox"/>	

B. Technical – General

1. Choose one of the following:		
a. This proposal is for a new script (set of characters):	<input checked="" type="checkbox"/>	
Proposed name of script:	<i>Glagolitic Extended</i>	
b. The proposal is for addition of character(s) to an existing block:	<input type="checkbox"/>	
Name of the existing block:		
2. Number of characters in proposal:	<i>38</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input checked="" type="checkbox"/>	B.2-Specialized (large collection) <input type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/>	G-Obscure or questionable usage symbols <input type="checkbox"/>	
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/>	
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/>	
b. Are the character shapes attached in a legible form suitable for review?	<input checked="" type="checkbox"/>	
5. Fonts related:		
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>Aleksandr Andreev (aleksandr.andreev@gmail.com; http://www.ponomar.net/cu_support.html)</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Menaion Unicode font licensed under GNU GPL by Aleksandr Andreev and Yuri Shardt</i>	
6. References:		
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/>	
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/>	
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)?	<input type="checkbox"/>	

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? If YES explain	<i>Submitted as L2/14-087; revised based on comments from the UTC</i>	YES
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.)? If YES, with whom? If YES, available relevant documents:	<i>Institut für Slavistik at the University of Vienna, Udmurt State University</i> <i>E-mail correspondence</i>	YES
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>Used by academics studying Glagolitic manuscripts. See Section 1, Introduction</i>	YES
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>Rare. See Section 1, Introduction</i>	YES
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Academic publications and digital storage of manuscripts. See Section 1</i>	YES
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:	<i>N/A</i>	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>N/A</i>	NO
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>N/A</i>	NO
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? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>N/A</i>	NO
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference:	<i>N/A</i> <i>N/A</i> <i>N/A</i>	NO
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)	<i>N/A</i>	NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:	<i>N/A</i>	NO