

Proposal to Encode a Slavonic Punctuation Mark in Unicode

Aleksandr Andreev* Yuri Shardt Nikita Simmons

PONOMAR PROJECT

1 Introduction

This document is a proposal to encode an additional character in the Supplemental Punctuation block of the Unicode Standard. This character is used in Church Slavonic manuscripts of the Ustav (Uncial) era for punctuation. The encoding of this character is required for two purposes: first, for the correct digital storage of the texts of these manuscripts and their search and display in electronic media; second, for use by palæographers, linguists, and scholars of liturgics in academic literature. The proposed character is summarized below. The examples demonstrate that this character can be encountered in a variety of printed or electronic texts that reproduce the contents of manuscripts.

2 Description of Character

The character is called the **Slavonic Spear** (Russian: *копие* (*kopie*)). It is used in certain Ustav manuscripts to indicate either a full stop, a medium stop, or the conclusion of a section. In the Ostromir Gospel, an important 11th Century manuscript, the Spear is used to indicate the end of a pericope or the end of a heading preceding a pericope; the Spear here commonly occurs next to a Middle Dot (period). We show this in Figure 1. The Spear plays a similar function in the Sava's Book,¹ another 11th Century manuscript, as can be seen from Figure 2. Both texts have been digitized and made available on-line by Manuscripts.ru; however, the project uses a custom font and an *ad hoc* codepage which encodes all Slavonic characters (including the Spear) in the Private Use Area of Unicode.

The usage of the Spear is not limited to the Cyrillic alphabet; it may also occur in Glagolitic manuscripts. Figure 3 presents an example of the Spear in the *Psalterium Sinaiticum*, an 11th Century Glagolitic Psalter, where the Spear is used to indicate the end of a Psalm verse. In Figure 4, the same character is seen in an academic reprint of the text of the Sinai Psalter (in Cyrillic transcription). In the critical edition of this manuscript published by Sever'yanov (1922), the Spear has been reprinted both in the text and in various comments of the critical apparatus.

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

¹So-named after the priest Sava (Sabbas), who inscribed his name on the original folios.

The character bears a visual similarity with several dash or hyphen characters already available in Unicode. However, while in all probability it has a similar historical origin as the dash or the Greek Paragraphos (U+2E0F), it has a different visual appearance, with an upturn on the left side of the character. All examples of the Spear that we have identified include this upturn. On the other hand, the Spear characters occurring in the sources are often different in length, as is characteristic of the manuscript tradition. Finally, since the Spear can be used both in Cyrillic and Glagolitic texts, it is most appropriately encoded in the Supplemental Punctuation block of Unicode, as proposed in Table 1 and the subsequent listing of character properties.

Table 1: Proposed Character

Character	Codepoint	Name
↵	U+2E43	SLAVONIC SPEAR

3 Character Properties

The following entry is proposed for addition to UnicodeData.txt:

```
2E43;SLAVONIC SPEAR;Po;0;ON;;;;;N;;;;;
```

References

- Scherkin, V. N. (1903). *Саввина книга*. Number v. 1, pt. 2 in Памятники старославянского языка. Изд. Отдѣленія русскаго языка и словесности Императорской академіи наук.
- Sever'yanov, S. N. (1922). *Синайская псалтырь: глаголическій памятник XI вѣка*. Памятники старославянского языка. Изд. Отдѣленія русскаго языка и словесности Россійской академіи наук.

Figure 1: Examples of the *Spear* used at the end of a pericope and at the end of a heading. Source: Ostromir Gospel.

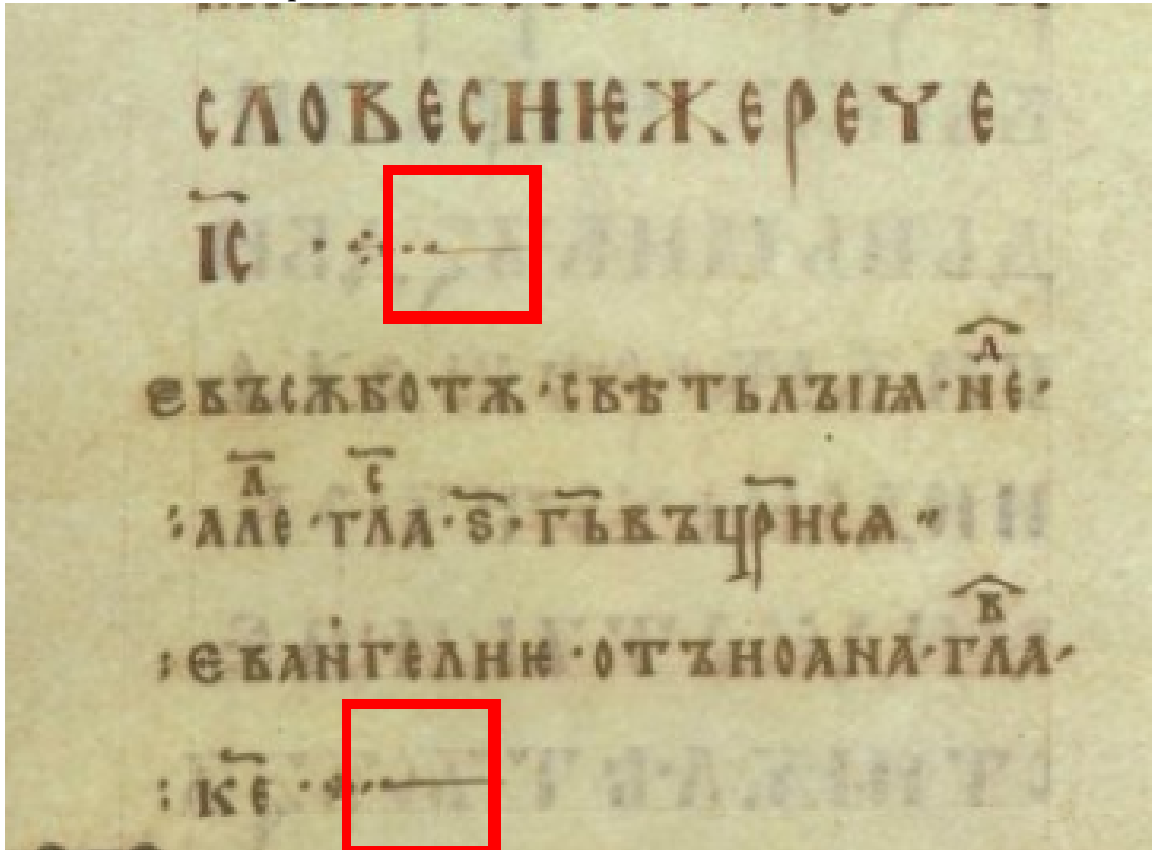


Figure 2: Examples of the *Spear* used at the end of a pericope and at the end of a heading. Source: Sava's Book as reproduced by Schepkin (1903).





НИКОВИ. ИДИ ШКО КЪРОКА. ДА БЖДЕ
 ТЪ ТИ. И ИСЦЪЛЪ СЛОУГА ЕГО ВЪ ТЪ ЧАСЪ. 
 ∴ СЖЪ Е ЕВА  ГЛА  Д.  Мо. IX.
 БЪ ОНО. МИМОИДЪ ИС. ВИДЪ ЧЛКА СЪДА
 ЦА НА МЪТЪНИЦИ. ИМЕНЬ МЪ МАДЕ
 А И ГЛА ЕМОУ ПО МИЕ ИДИ. И ВЪСТАВЪ

Figure 3: Examples of the *Spear* used at the end of psalm verses written in the Glagolitic script. Source: Psalterium Sinaiticum.

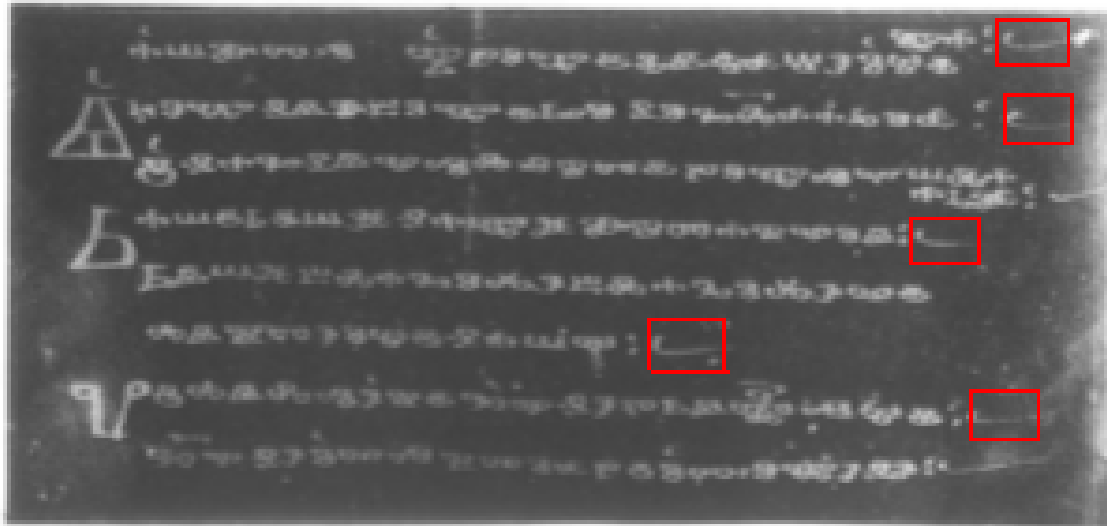


Figure 4: Examples of the *Spear* used at the end of a psalm verse. Source: Psalterium Sinaiticum as reproduced by Sever'yanov (1922).

на гѣ сна хѣ его: └─	1	скъдельника съкроуши:	
3 Растрѣгнѣмъ жъзы ихъ	2	ї нѣмѣ црѣ разоумѣте	10
ї отъкрѣзѣмъ отъ насъ	3	накажѣте сѣа въсі сѣдѣ-	
нго ихъ:	4	цеи земли: └─	
4 Живѣи на вѣхъ посмѣетъ	5	Работанте гѣи съ страхомъ:	11
сѣа емоу і гѣ поржгаетъ сѣа ^{мѣ} :	6	и радоуѣте сѣа емоу съ тре-	
5 Тъгда възглѣтѣ къ нимъ	7	петомъ: └─ └─	
гнѣвомъ своимъ: и ѣро-	8	Принѣте наказанѣ: еда	12
стыж своимъ съмиатѣтъ ^{мѣ} └─	9	когда прогнѣваетъ сѣа	

**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 a Slavonic Punctuation Mark in Unicode	
2. Requester's name:	<i>Aleksandr Andreev, Yuri Shardt, and Nikita Simmons</i>	
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>	
4. Submission date:	<i>07/08/2013</i>	
5. Requester's reference (if applicable):	<i>N/A</i>	
6. Choose one of the following:		
This is a complete proposal:	YES	
(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):	NO	
Proposed name of script:		
b. The proposal is for addition of character(s) to an existing block:	YES	
Name of the existing block:	<i>Supplemental Punctuation</i>	
2. Number of characters in proposal:	1	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary	B.1-Specialized (small collection)	<input checked="" type="checkbox"/> 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?	<i>Aleksandr Andreev</i>	
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>Hirmos Ponomar font distributed by Aleksandr Andreev, Yuri Shardt, Nikita Simmons under GNU GPL http://www.ponomar.net/ or aleksandr.andreev@gmail.com</i>	
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: N4102-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		NO
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>Slavonic Typography Society</i> <i>Online discussion at http://cslav.orthonet.ru/</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:		NO
4. The context of use for the proposed characters (type of use; common or rare) Reference:		Rare
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>Academic literature on Slavonic paleography</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:		NO
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		N/A
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:		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:		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>Section 2., description of character</i>	YES YES
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:		NO N/A
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)		NO
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		NO