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 widely used in Church Slavonic manuscripts, written both in the Cyrillic and Glagolitic scripts, for punctuation. It may also be encountered in academic texts that study or reproduce manuscripts. The encoding of this character is thus required for two purposes: first, for the correct digital storage of the contents of 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.

2   Description of Character

We propose to name the character Slavonic Paragraphos.[1] The Paragraphos is a symbol that is found in Greek manuscripts and can serve a variety of functions, in particular, to indicate a marginal note or the beginning or ending of a sentence. The proposed Slavonic Paragraphos is used in certain Ustav (Uncial) manuscripts to indicate either a full stop, a medium stop, or the conclusion of a section. For example, in the Ostromir Gospel, an important 11th Century manuscript, this symbol is used to indicate the end of a pericope or the end of a heading preceding a pericope; in this usage, it commonly occurs next to a Middle Dot (period) as can be seen in Figure 1. The Paragraphos plays a similar function in the Sava's Book,[2] another 11th Century manuscript, as can be seen from Figure 2. Both of these texts – the Ostromir Gospel and the Sava's Book – have been digitized and made available on-line by Manuscripts.ru; however, that project uses a custom font and an *ad hoc* codepage which encodes all Slavonic characters (including the Paragraphos) in the Private Use Area of Unicode.

---

[*]Corresponding author: `aleksandr.andreev@gmail.com`.

[1]This character was originally proposed with the name "Slavonic Spear". The name "Spear" is a translation of the Slavonic term "Kopie", first used for this character by Cleminson et al. (2010). However, consultation with the authors of that paper revealed that the term "Kopie" was created as a translation of the Greek term obelus. In modern typography, the term obelus most commonly refers to the division sign and related symbols used for obelism. We believe that the present character is more properly associated with the paragraphos and not the obelus, based both on its appearance and function. Thus we now propose a different name.

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

The usage of the Paragraphos is not limited to the Cyrillic script; it may also occur in Glagolitic manuscripts. Figure 3 presents an example of this character in the *Psalterium Sinaiticum*, an 11th Century Glagolitic Psalter, where the Paragraphos 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 Paragraphos has been reprinted both in the text and in various comments of the critical apparatus.

Finally, in Figure 5 we present an example of the Slavonic Paragraphos used in a modern setting to separate blocks of lyrics in Znamenny musical notation. Here, the character is used out of a desire to imitate ancient manuscripts.

The Slavonic Paragraphos character bears a visual similarity with, and has a similar functional purpose as, several dash or hyphen characters already available in Unicode. As well, the Unicode Standard includes already the Greek Paragraphos (U+2E0F) and two related characters, the Forked Paragraphos and the Reversed Forked Paragraphos. However, the Slavonic Paragraphos has a somewhat different visual appearance than these characters, with an upturn on the left side of the character (all examples of the Slavonic Paragraphos that we have identified include this upturn). As well, while the three Greek Paragraphoi encoded in Unicode typically occur along the lower baseline, the Slavonic Paragraphos is typically raised from the baseline. Therefore, it would not be appropriate to use the codepoint for the Paragraphos (U+2E0F) or its variants to encode the Slavonic Paragraphos. Since the Slavonic Paragraphos can be used both in Cyrillic and Glagolitic texts, it is most appropriately encoded at a new codepoint 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 PARAGRAPHOS

3   Character Properties

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

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

Above, we described the primary function of the Slavonic Paragraphos as indicating the end of a sentence, phrase or pericope. In this usage, line breaking commonly occurs following the Paragraphos. In addition, the Slavonic Paragraphos is usually preceded by one or more decorative punctuation characters, such as a colon, middle dot, or three dot punctuation. These characters are never separated from the Slavonic Paragraphos by a line break, as can be seen in Figures 1, 3 and 5. It is thus proposed that the line breaking property of the Slavonic Paragraphos be set to Break After (BA). This is also consistent with the line breaking property of the various Greek Paragraphoi already in Unicode. Therefore, the following entry is proposed for addition to LineBreak.txt:

```
2E43;BA # SLAVONIC PARAGRAPHOS
```

Finally, because this character is used both in the Cyrillic and Glagolitic scripts, it is proposed that the character's script property be set to Common, by adding the following entry to Scripts.txt:

```
2E43          ; Common # Po    SLAVONIC PARAGRAPHOS
```

This would agree with the script property of other characters in the Supplemental Punctuation block.

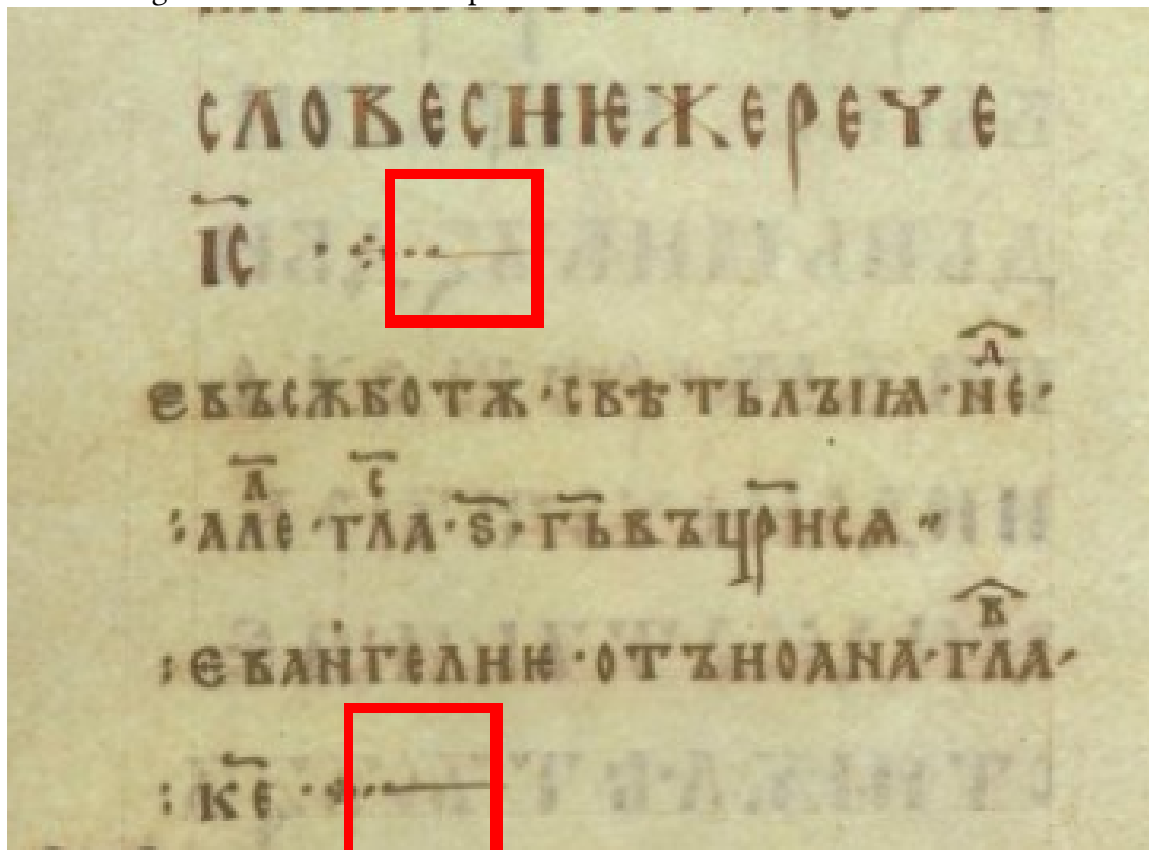Figure 1: Examples of the Slavonic Paragraphos used at the end of a pericope and at the end of a heading. Source: Ostromir Gospel.

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

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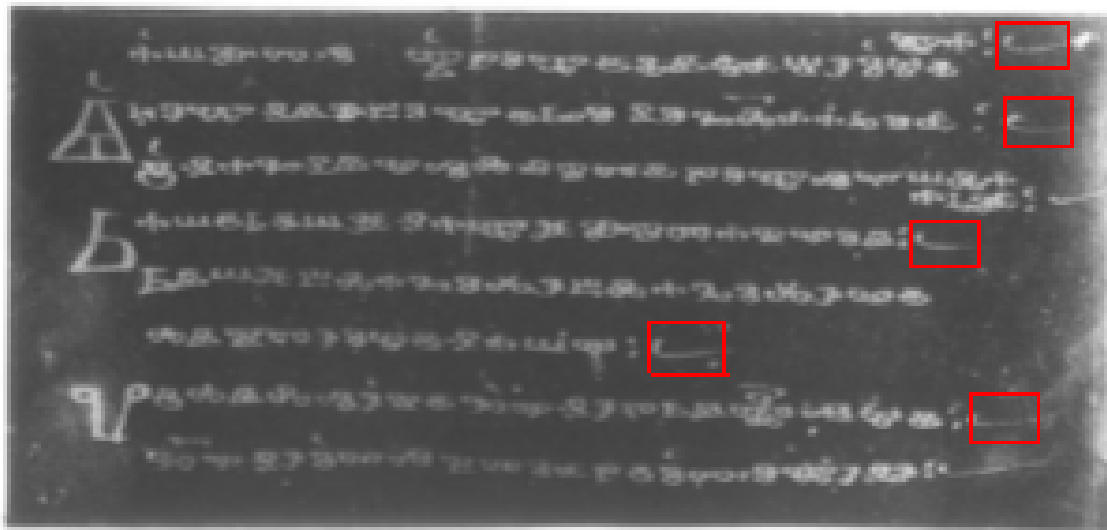

Figure 4: Examples of the Slavonic Paragraphos used at the end of a psalm verse. Source: Psalterium Sinaiticum as reproduced by Sever'yanov (1922).

Figure 5: Examples of the Slavonic Paragraphos used to separate lyrics in Znamenny musical notation. Source: Materials of the 2nd Conference of Practitioners of Znamenny Chant, Moscow, 2004.

References

Cleminson, R., V. Baranov, A. Rabus, D. Birnbaum, and H. Miklas (2010). Proposal for a unified encoding of Early Cyrillic glyphs in the Unicode Private Use Area. *Scripta & e-Scripta* (8-9), 9–26.

Schepkin, 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.

A. Administrative

1. **Title:** *Proposal to Encode a Slavonic Punctuation Mark in Unicode*
2. Requester's name: *Aleksandr Andreev, Yuri Shardt, and Nikita Simmons*
3. Requester type (Member body/Liaison/Individual contribution): *Individual contribution*
4. Submission date: *09/23/2013*
5. Requester's reference (if applicable): *N/A*
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: *Supplemental Punctuation*
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)   X   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
   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 distributed by Aleksandr Andreev, Yuri Shardt, Nikita Simmons under GNU GPL
      http://www.ponomar.net/ or aleksandr.andreev@gmail.com
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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?   *YES*
   If YES explain   *This is a revision of L2/13-140, based on feedback from UTC and discussion with experts*
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?   *Slavonic Typography Society; Victor A. Baranov; Ralph Cleminson*
      If YES, available relevant documents: *E-mail correspondence*
3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included?   *NO*
   Reference:
4. The context of use for the proposed characters (type of use; common or rare)   *Rare*
   Reference:
5. Are the proposed characters in current use by the user community?   *YES*
      If YES, where? Reference: *Academic literature on Slavonic paleography; Znamenny sheet music*
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely
   in the BMP?   *NO*
        If YES, is a rationale provided?
        If YES, reference:
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?   *NO*
        If YES, is a rationale for its inclusion provided?
        If YES, reference:
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:
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: *Section 2., description of character*
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:
       Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?   *N/A*
         If YES, reference:
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)

13. Does the proposal contain any Ideographic compatibility characters?   *NO*
       If YES, are the equivalent corresponding unified ideographic characters identified?
         If YES, reference: