PONOMAR PROJECT

Proposal to Encode An Outstanding Early Cyrillic Character in Unicode

Aleksandr Andreev, Yuri Shardt, Nikita Simmons

In early Cyrillic printed editions and manuscripts one finds many combining (superscript) characters. The majority of these have been encoded in the Cyrillic Extended A block with a number of additional characters in the Cyrillic Extended B block. Notably, one combining Cyrillic letter is missing—the letter *ef* (ф). Based on an examination of the literature, it can be seen that such a combining character actually exists. Thus, we propose for encoding one additional character, Combining Cyrillic Letter Ef. Since these symbols are often found in Slavonic Church books, this symbol should be encoded at the last available spot U+A69E in the Extended Cyrillic Block B in the Unicode standard.

New Proposed Character

Table 1 presents a summary of the proposed new character for encoding in Unicode: COMBINING CYRILLIC LETTER EF, which is commonly found in many early Slavonic printed books, including the monumental Ostrog Bible. This character can be found with or without a pokryttie and is used to denote an abbreviation. Its DUCET collation order should be given as "ф << $\overset{\text{ф}}{\circ}$ < x." Examples of these characters can be found in Figure 1 and Figure 2.[1]
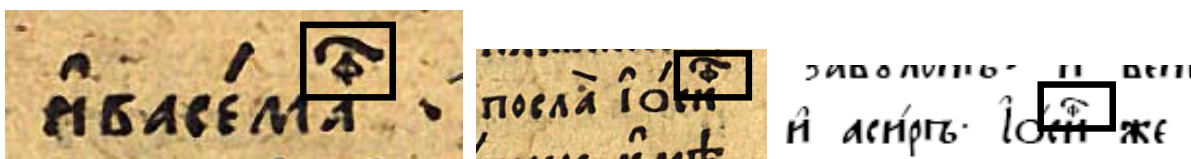

Figure 1: Extracts from the Ostrog Bible showing a superscript ef (boxed) (left) Genesis 36:4, (centre) Genesis 46:5, and (right) Exodus 1:4. The first two examples are from the actual Ostrog Bible printed editions, while the third example is from a digitisation and scholarly analysis of the Ostrog Bible.
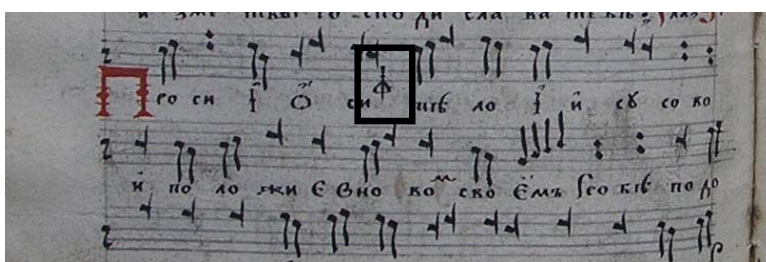

Figure 2: Extract from Lenten and Flowery Triodion set in Kievan notation of early 18th century provenance. It can be noted that in this manuscript, the superscript letters are not placed over a letter, but between the two characters (or space).

---

[1] Assistance from Messrs. Konstantin Spektorov and Aleksandr Kovalenin is acknowledged in uncovering examples of combining ef in the print editions.

Table 1: Summary of the Proposed Characters for Encoding

Proposed Character	Proposed Name	Location	Comments
ф	COMBINING CYRILLIC LETTER EF	U + A69E	

ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646[1]
Please fill all the sections A, B and C below.
Please read Principles and Procedures Document (P & P) from http://www.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://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html**.**
See also http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html **for latest *Roadmaps*.**

A. Administrative

1. **Title:** *Proposal to Encode Some Outstanding Early Cyrillic Characters in Unicode*
2. Requester's name: *, Aleksandr Andreev, Yuri Shardt, Nikita Simmons*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *January 21st, 2013*
5. Requester's reference (if applicable):
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: *U+ A69E*
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) *Yes* 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. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for
    publishing the standard? *Yuri Shardt*
    If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools
    used: *Hirmos Ponomar v.14 (contact Yuri Shardt at yuri.shardt@ualberta.ca for the font)*
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 http://www.unicode.org/Public/UNIDATA/UCD.html 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? — *NO*
   If YES explain

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? — *Academics; Liturgical Use*
   If YES, available relevant documents:

3. Information on the user community for the proposed characters (for example:
   size, demographics, information technology use, or publishing use) is included? — *small, but active*

   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: — *In typesetting older documents authentically*

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?
   If YES, reference:

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:

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? — *NO*
   If YES, reference:

10. Can any of the proposed character(s) be considered to be similar (in appearance or function)
    to an existing character? — *NO*
    If YES, is a rationale for its inclusion provided? — *NO*
    If YES, reference:

11. Does the proposal include use of combining characters and/or use of composite sequences? — *YES*
    If YES, is a rationale for such use provided? — *YES*
    If YES, reference:
    Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?
    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 character(s)? — *NO*
    If YES, is the equivalent corresponding unified ideographic character(s) identified?
    If YES, reference: