ISO/IEC JTC1/SC2/WG2

Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation
Международная организация по стандартизации

Doc Type: **Working Group Document**
Title: **Proposal to add a character used in printed Glagolitic works to the UCS**
Source: **Zoran Vukojević**
Status: **Individual Contribution**
Date: **2019-07-20**

This proposal requests the encoding of a variant of the Glagolitic letter červ. If this proposal is accepted, the following characters will exist:

Ɐ   2C2F   GLAGOLITIC CAPITAL LETTER CHERV
• used as an alternative to Ⱝ 2C1D Glagolitic capital letter chrivi, and used combined with 0303 õ combining tilde or with 0483 ṑ combining Cyrillic titlo as an et cetera sign.

ⱟ   2C5F   GLAGOLITIC SMALL LETTER CHERV
• used as an alternative to ⱝ 2C4D Glagolitic small letter chrivi, and used combined with 0303 õ combining tilde or with 0483 ṑ combining Cyrillic titlo as an et cetera sign.

The character is used in works printed in the Glagolitic alphabet. The first dated printed book in Glagolitic is the Missal from 1483. Religious works in the Glagolitic alphabet continued to be printed until 1927 when the last Missal in the Croatian recension of Church Slavonic was printed for the use of churches in Croatia. Scholarly works on and in the Glagolitic alphabet are still being produced.

The letter červ was used in two versions, the first is already encoded in Unicode as ✿ 2C1D GLAGOLITIC CAPITAL LETTER CHRIVI and as ✿ 2C4D GLAGOLITIC SMALL LETTER CHRIVI. The second is the one proposed in this document, Ɐ GLAGOLITIC CAPITAL LETTER CHERV and ⱟ GLAGOLITIC SMALL LETTER CHERV. It is derived from the handwritten form of the letter. Although both variants were used to write the sound /tʃ/, the second one was used with a combining tilde or combining titlo as an *et cetera* sign. When a combining tilde or combining titlo is used with the first variant, it indicates either the numeral 1000 – refer to figure 3 where it is used in writing the year 1561 – or as an indicator of an abbreviation – refer to figure 2.

Unicode already encodes variant letter forms for Glagolitic A – Azu and Trokutasti A. Two variant letter forms for Glagolitic M are also encoded – Myslite and Latinate Myslite. It is necessary to encode the Glagolitic letter Cherv, to indicate correctly the *et cetera* sign, and to show the palaeographic variations in books between words printed with Cherv and those printed with Chrivi – refer to figure 4 for an example. The two variants cannot be unified.

As for the name Cherv, it is the name of the letter that was used during the time when the printed works were published. Refer to figure 6 for evidence of the use of the name.

1. Unicode Character Properties. Character properties are proposed here.

2C2F; GLAGOLITIC CAPITAL LETTER CHERV;Lu;0;L;;;;;N;;;;;
2C5F; GLAGOLITIC SMALL LETTER CHERV;Lu;0;L;;;;;N;;;;;

2. Bibliography

Bukvar slavenskij 1753. *Bukvarъ slavenskiĵ pismeny veličaĵšagō učitelê B. Ieronima stridonskago napečatan*. Rome: Sveti Sobor ot razmnoženiê very [Sacred Congregation for the Propagation of the Faith].

Dalmatin, A. and Istrijan, S. 1561. *Tabla za dicu: Edna malahna knižice iz koih se ta mlada predraga ditca, tere priprosti ljudi zglagolskimi slovima čtati, i poglavitei, i potribnei Artikuli, ili členi Ove prave stare Karstianske vere, koĵa svakoga čovika izveliča lahko mogu naučiti. ABECEDARIVM, Un der gancze Cathechiſmus / one außlegung / in der Chrobatiſchen Sprach*. Tubingen.

Levaković, Rafail. 1629. *Azbukividnêk slovinskiĵ, iže općenim načinom Psalterić nazivaet se*. Rome: Sveta Skupšćina ot razmnoženiê Veri [Sacred Congregation for the Propagation of the Faith].

3. Figures.

Figure 1. Example from Levaković. 1629:15 showing the ɥ GLAGOLITIC SMALL LETTER CHERV combined with ◌̃ COMBINING TILDE, used as an et cetera sign (indicated by the blue arrows). Indicated by the red arrows is the ✿ GLAGOLITIC SMALL LETTER CHRIVI in its usual use for the sound /tʃ/.

Figure 2. Example from Leváković. 1629:25 showing the ⛨ GLAGOLITIC SMALL LETTER CHRIVI combined with ◌̃ COMBINING TILDE, used in different abbreviations. The abbreviations, starting in the tenth row, are Mčk for mučenik (martyr, masculine, nominative case), Mčka for mučenika (martyr, masculine, genitive case), Mčca for mučenica (martyr, feminine, nominative case).
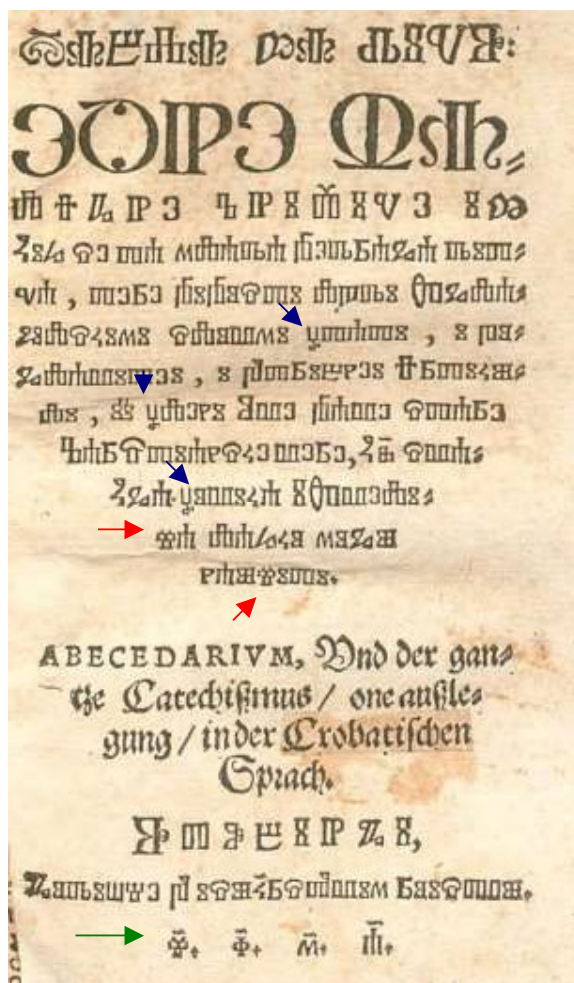
Figure 3. Dalmatin and Istrijan. 1561: title page showing the use of ꙏ GLAGOLITIC SMALL LETTER CHERV indicated by blue arrows, and the use of ꙮ GLAGOLITIC SMALL LETTER CHRIVI, indicated by red arrows, for the sound /tʃ/. The numeral 1000 is denoted by ꙮ GLAGOLITIC SMALL LETTER CHRIVI combined with ̃ COMBINING TILDE, indicated by the green arrow.
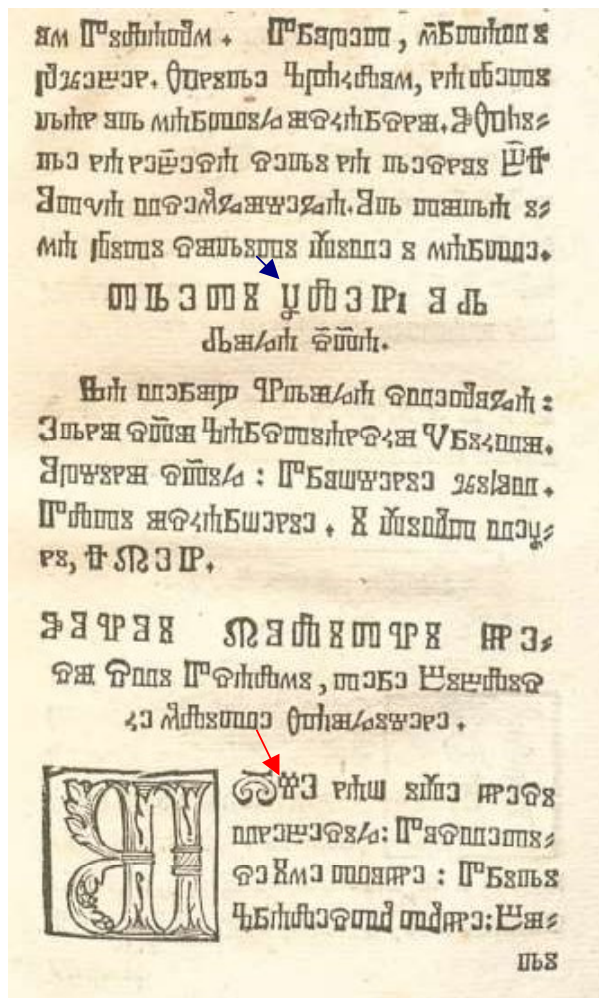
Figure 4. Dalmatin and Istrijan. 1561: title page showing the use of Ⱍ GLAGOLITIC CAPITAL LETTER CHERV for the sound /tʃ/ indicated by the blue arrow. Ⱚ GLAGOLITIC CAPITAL LETTER CHRIVI is also used on the same page for the sound /tʃ/, indicated by the red arrow.
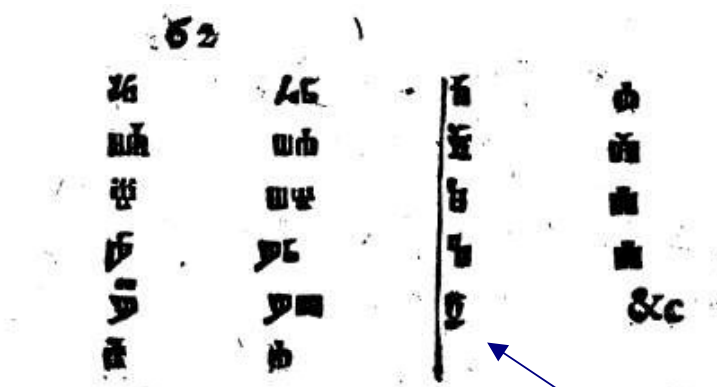

Figure 5. Bukvar slavenskij 1753:62 with a table of ligatures and abbreviations used in Glagolitic. The last item in the right hand column is Ⱍ GLAGOLITIC SMALL LETTER CHERV combined with ◌̃ COMBINING TILDE. This is Ⱍ̃ the Glagolitic *et cetera* sign.

Figure 6. *Bukvar slavenskij* 1753:6 showing the name used in Croatia for the letter was Cerv, pronounced as červ /tʃerv/.

A. Administrative
1. Title
Proposal to add a character used in printed Glagolitic works to the UCS
2. Requester's name
Zoran Vukojević.
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution.
4. Submission date
2019-07-20
5. Requester's reference (if applicable)
6. Choose one of the following:
6a. This is a complete proposal
Yes.
6b. More information will be provided later
No.

B. Technical – General
1. Choose one of the following:
1a. This proposal is for a new script (set of characters)
No.
1b. Proposed name of script
1c. The proposal is for addition of character(s) to an existing block
Yes
1d. Name of the existing block
Glagolitic
2. Number of characters in proposal
2.

3. Proposed category (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)
Category B.1.
4a. Is a repertoire including character names provided?
Yes.
4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes.
4c. Are the character shapes attached in a legible form suitable for review?
Yes.
5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?
Zoran Vukojević.
5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:
Zoran Vukojević, Font Forge.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes.
7. 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)?
Yes.
8. 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/ Public/UNIDATA/UnicodeCharacterDatabase.html and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.
See above.

C. Technical – Justification
1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
No.
2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
No.
2b. If YES, with whom?
2c. 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?
Scholarly, ecclesiastical, and library communities.
4a. The context of use for the proposed characters (type of use; common or rare)
Originally used to write Old Church Slavonic, and used in parts of medieval and early modern Croatia for both secular and ecclesiastical purposes. Also used to write old and modern Croatian.
4b. Reference
5a. Are the proposed characters in current use by the user community?
Yes.
5b. If YES, where?
By scholars, churchmen, librarians, and Glagolitic enthusiasts.
6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?
Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Accordance with the Roadmap. Keep with other Glagolitic characters.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes, they should be encoded in a single block as presented here.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

No.

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

11e. If YES, reference

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

14c. If YES, reference