PONOMAR PROJECT

Proposal to Encode Chinese Characters Used for Transliterating Slavonic

Yuri Shardt, Mitrophan Chin, Aleksandr Andreev

Background

The development of renewed contacts between the Russian and Chinese Empires in the time period between A.D. 1700 and 1900 led to an interest in the translation of Slavonic literary documents into both classical and vernacular Chinese. One of the main forces encouraging these translations was the Russian Mission (roughly equivalent to an embassy) in Beijing. Many of these translations are unusual in that they created new Hanzi characters to represent syllables not usually found in Chinese. In order to properly digitise these documents, there is a need to encode these characters in Unicode, so that their significance, usage, and occurrence can be more easily elucidated.

One area where there was an acute need for special Hanzi characters was in the translation of liturgical texts into Chinese. In order to accomplish this task, Archimandrite Gurias (*né* Grigory Platonovich Karpov)[1], head of the 14th Russian Mission in Beijing, devised a series of Hanzi characters for representing syllable structures not found in Chinese. Subsequently, under the direction of Archimandrite Innocent (*né* Ivan Appolonovich Figurovsky)[2], the head of the 18th Russian Mission, a decision was taken to translate the texts into the vernacular Chinese language, primarily Mandarin, and to simplify the characters used. Nevertheless, a few new characters were introduced.

Table 1 shows a summary of the proposed 18 characters along with their standard Mincho (print) and Kaishu (brush stroke) forms, names, and additional information. Of the proposed characters, all but the abbreviation for Christ and the simplified *ri* form were first proposed by Archimandrite Gurias. The increased emphasis on translating into the vernacular Chinese language led to character simplification. In most cases, the right-hand component was dropped and the resulting character was used. The following cases have special simplifications, namely □雷爾simplified to 列, □魯爾to 嚕, □伊克to 乞, □伊合 to 吸, □耶格 to 碣, □耶克to either楷 or 碣, □耶合to叶, and □伊格to吉. All of the simplified forms are already encoded in Unicode and hence do not need to be added. In addition to the simplification, Archimandrite

---

[1] Later, he became the Archbishop of Tauria and Simferopol (1867). In 2008, he was glorified by the Ukrainian Orthodox Church. His name is also given as Gury, which is a transliteration based on the Slavonic form.
[2] Later, he became the Bishop of Pereslavl (1902), the Archbishop of Beijing (1921), and then Metropolitan of Beijing.

Innocent introduced an abbreviation character for the word *Christ*, which was then used in the translations (see for example, Figure 19). Finally, it can be noted that the simplified *ri* form was also developed and used. It is proposed that these glyphs be encoded as a block and be sorted based on their underlying Slavonic phonetic values and sort order.

Table 1: Summary of proposed new characters

Name	Composition	Code Point	Form		Comments
			Kaishu	Mincho	
CJK SLAVONIC IDEOGRAPH VIN	英微 82F1,5FAE		英微	薇	equivalent to *vin* (вин) as in Навин (Nun, 那英微)
CJK SLAVONIC IDEOGRAPH GE	耶格 8036,683C		耶格	聯格	equivalent to *ge* (ге) as in Нигер (Niger, 尼耶格爾)
CJK SLAVONIC IDEOGRAPH GI	伊格 4F0A,683C		伊格	俕格	equivalent to *gi* (ги) as in Сергия (Sergius, 些爾伊格乙)
CJK SLAVONIC IDEOGRAPH KE	耶克 8036,514B		聎	聎	equivalent to *ke* (ке) as in Кесария (Caesarea, 耶克薩利爾亞)
CJK SLAVONIC IDEOGRAPH KI	伊克 4F0A,514B		俔	俔	equivalent to *ki* (ки) as in Езекия (Hezekiah, 耶捷伊克亞)
CJK SLAVONIC IDEOGRAPH RA	拉爾 62C9,723E		攊	攊	equivalent to *ra* (ра) as in Израиль (Israel 伊斯拉爾伊利)

Name	Composition	Code Point	Form		Comments
			Kaishu	Mincho	
CJK SLAVONIC IDEOGRAPH RAN	□郎爾 90CE,723E		郎爾	郎爾	equivalent to *ran* (ран) as in Аран (Haran, 哈爾□郎爾)
CJK SLAVONIC IDEOGRAPH RE	□列爾 5217,723E		列爾	列爾	equivalent to *re* (ре) as in Назарет (Nazareth, 那匝□列爾特)
CJK SLAVONIC IDEOGRAPH REN	□楞爾 695E,723E		楞爾	楞爾	equivalent to *ren* (рен) as in Терентий (Terence, 鐵□楞爾提乙)
CJK SLAVONIC IDEOGRAPH REIA	□雷爾 96F7,723E		雷爾	雷爾	equivalent to *reia* (рея) as in назареянин (Nazarene, 那作□雷爾)
CJK SLAVONIC IDEOGRAPH RI	□利爾 5229, 723E		稱爾	稱爾	equivalent to *ri* (ри) as in Христос (Christ, 合□利爾斯托斯)
CJK SLAVONIC IDEOGRAPH SIMPLIFIED RI	□利尔 5229,5C14		称	称	Simplified variant of □利爾.
CJK SLAVONIC IDEOGRAPH RIN	□凌爾 51CC,723E		凌爾	凌爾	equivalent to *rin* (рин) as in Коринф (Corinth, 适□凌爾福)

Name	Composition	Code Point	Form		Comments
			Kaishu	Mincho	
CJK SLAVONIC IDEOGRAPH RO	□羅爾 7F85,723E		羅爾	羅爾	equivalent to *ro* (ро) as in романский (Roman, □羅·爾瑪)
CJK SLAVONIC IDEOGRAPH RON	□隆爾 9686,723E		隆爾	隆爾	equivalent to *ron* (рон) as in Аарон (Aaron, 阿阿□·隆爾)
CJK SLAVONIC IDEOGRAPH RU	□魯爾 9B6F,723E		魯爾	魯爾	equivalent to *ru* (ру) as in Иерусалим (Jerusalem, 耶·□魯爾薩利木)
CJK SLAVONIC IDEOGRAPH KHE	□耶合 8036,5408		耶合	耶合	equivalent to *khe* (хе) as in Сихем (Shechem, 西□耶·合木)
CJK SLAVONIC IDEOGRAPH KHI	□伊合 4F0A,5408		伊合	伊合	equivalent to *khi* (хи) as in Мелхий (Melchi, 芈利□·伊合乙)
CJK SLAVONIC IDEOGRAPH KHRISTOS	日合一 5408,4E00		合	合	abbreviation for the word *Christ* (Христос, Khristos). The full form in Classical Chinese would be 合□利爾·斯托斯.
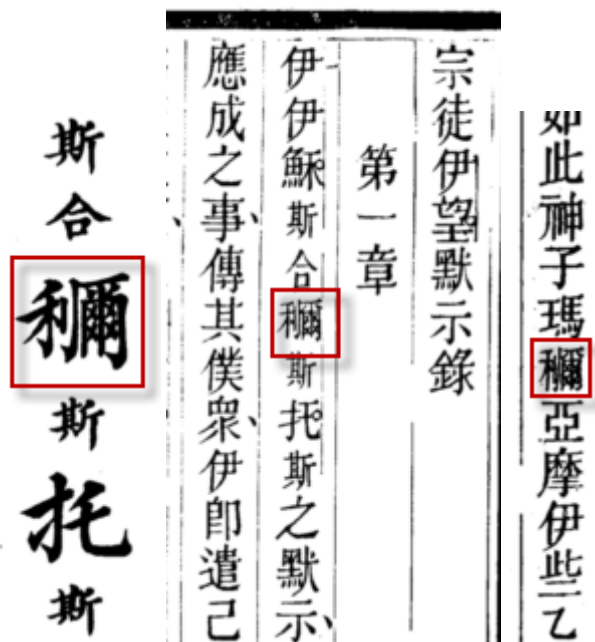
Figure 1: Example of CJK SLAVONIC IDEOGRAPH RI (left) from the cover page of the New Testament translated by Archimandrite Gurias (Archimandrite Gurias, 1864), (middle) Book of Revelation showing use of the same character (Archimandrite Gurias, 1864), and from the front matter from the New Testament (Archimandrite Gurias, 1864).

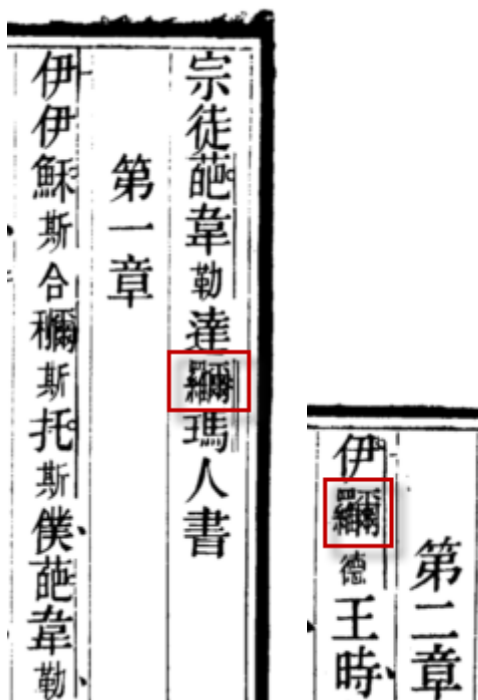Figure 2: Example of CJK SLAVONIC IDEOGRAPH SIMPLIFIED RI from the front matter from the New Testament (Archimandrite Gurias, 1864) in chopmark seal style.

Figure 3: Example of CJK SLAVONIC IDEOGRAPH RO (left) from the start of the Epistle to the Romans (Archimandrite Gurias, 1864) and the name Herod at Matthew 2:1 from the New Testament (Archimandrite Gurias, 1864).
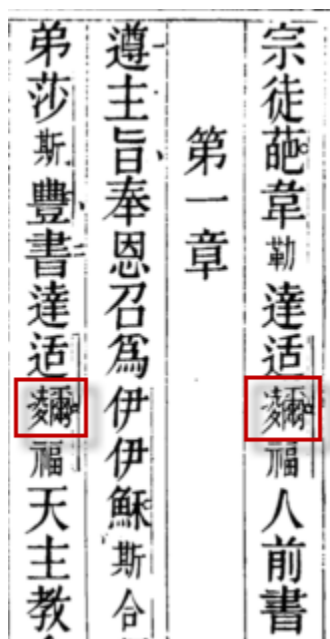
Figure 4: Example of CJK SLAVONIC IDEOGRAPH RIN from the start of the First Epistle to the Corinthians (Archimandrite Gurias, 1864).
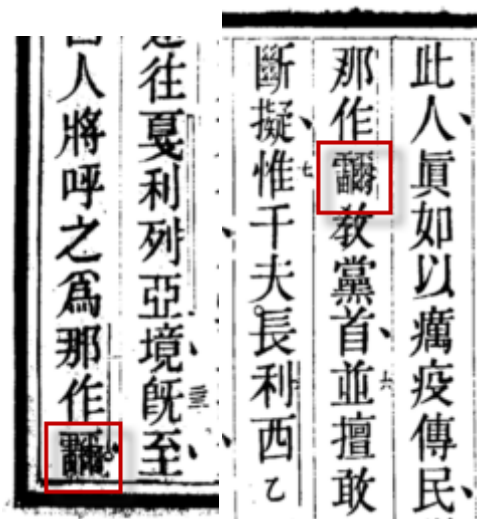
此人眞如以癩疫傳民、
斷擬惟千夫長利西
那作蠶教黨首並擅敢
往夏利列亞境旣至
一人將呼之爲那作蠶

Figure 5: Example of CJK SLAVONIC IDEOGRAPH REIA from (right) Matthew 2:23 and (left) Acts 24:5 (Archimandrite Gurias, 1864).

時伊伊穌斯旣誕於伊屋
自東方至耶蠶薩利
日
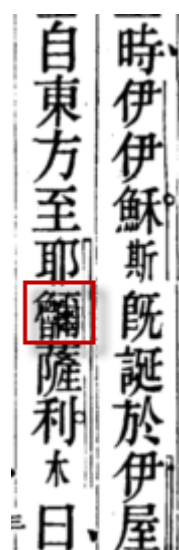
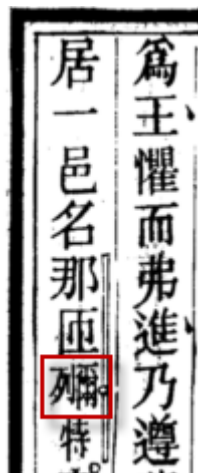Figure 6: Example of CJK SLAVONIC IDEOGRAPH RU from Matthew 2:1 (Archimandrite Gurias, 1864).

Figure 7: Example of CJK SLAVONIC IDEOGRAPH RE from Matthew 2:23 (Archimandrite Gurias, 1864).
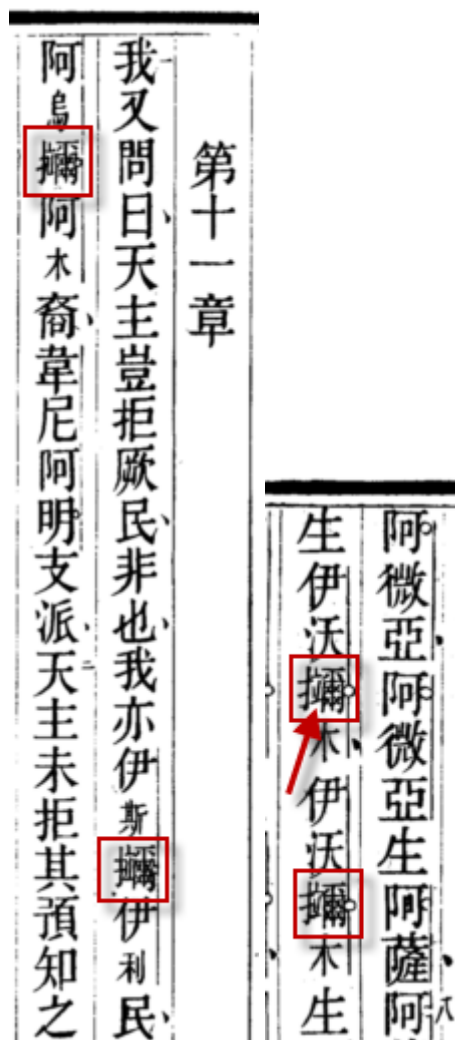
Figure 8: Example of CJK SLAVONIC IDEOGRAPH RA (left) Romans 11:1 and (right) Matthew 1:8 (Archimandrite Gurias, 1864).

Figure 9: Example of CJK SLAVONIC IDEOGRAPH KI from Matthew 1:9-10 (Archimandrite Gurias, 1864).

Figure 10: Example of CJK SLAVONIC IDEOGRAPH KHI from Matthew 2:18 (Archimandrite Gurias, 1864).

Figure 11: Example of CJK SLAVONIC IDEOGRAPH GE from Mark 3:17 (Archimandrite Gurias, 1864).
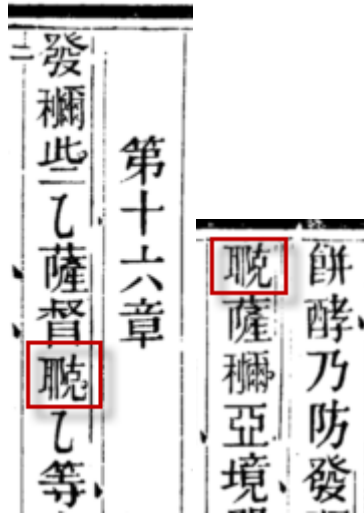
Figure 12: Example of CJK SLAVONIC IDEOGRAPH KE from (left) Matthew 16:1 and (right) Matthew 16:13 (Archimandrite Gurias, 1864).
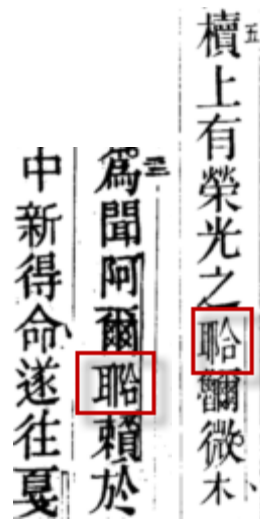

Figure 13: Example of CJK SLAVONIC IDEOGRAPH KHE from (left) Matthew 2:22 and (right) Hebrews 9:5 (Archimandrite Gurias, 1864).
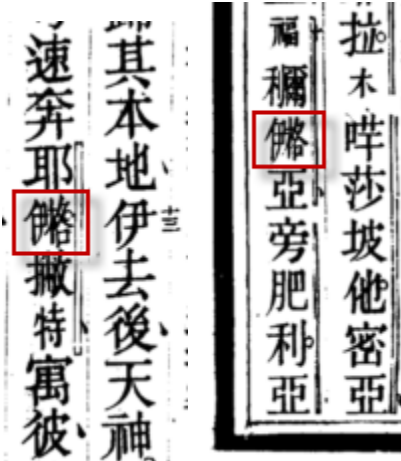
Figure 14: Example of CJK SLAVONIC IDEOGRAPH GI from (left) Matthew 2:13 and (right) Acts 2:10 (Archimandrite Gurias, 1864).
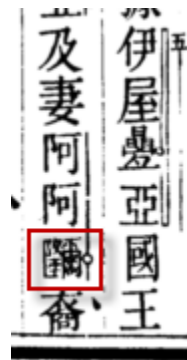

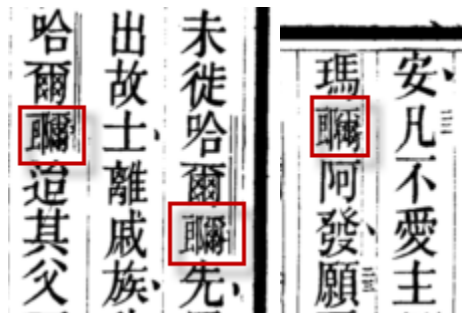Figure 15: Example of CJK SLAVONIC IDEOGRAPH RON from Luke 1:5 (Archimandrite Gurias, 1864).

Figure 16: Example of CJK SLAVONIC IDEOGRAPH RAN from (left) Acts 7:2, 4 and (right) 1 Corinthians 16:22 (Archimandrite Gurias, 1864).

Figure 17: Example of CJK SLAVONIC IDEOGRAPH VIN from Hebrews 4:8 (Archimandrite Gurias, 1864).

Figure 18: Example of CJK SLAVONIC IDEOGRAPH REN from (left) *The Summary of New Testament Salvation History: Life of Christ* translated by Archimandrite Gurias (Archimandrite Gurias, 1861) and (right) title page of the *Mirror of an Orthodox Confession* by St. Demetrius (Dimitry) of Rostov translated by Archimandrite Gurias (Saint Demetrius of Rostov, 1860).
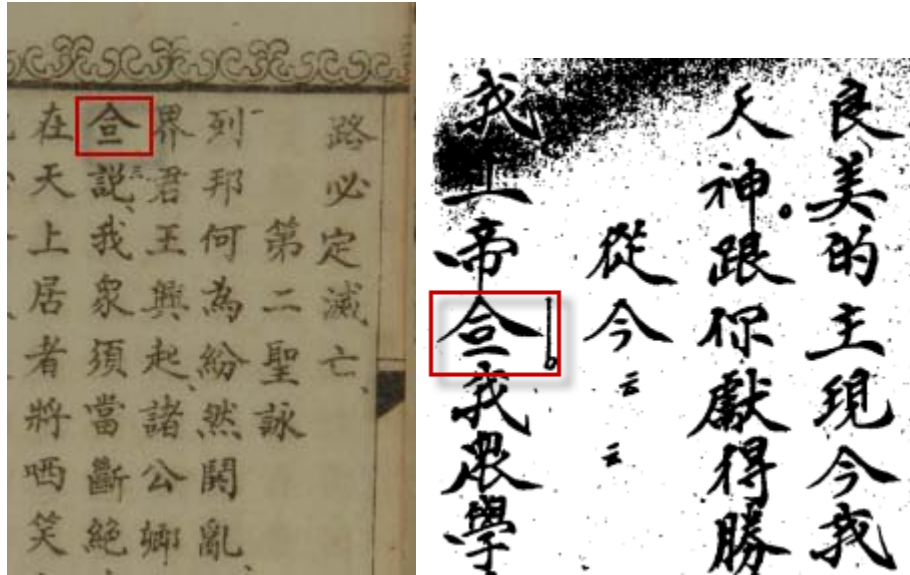
Figure 19: Example of CJK SLAVONIC IDEOGRAPH KHRISTOS from (left) Psalm 2:2 from Metropolitan Innocent's translation of the Psalter into Mandarin (Archimandrite Innocent, 1910) and (right) extract from a manuscript of the Horologion translated into the vernacular Chinese language found in Harbin, China from 1930/40.

References

New Testament in Chinese. (1864). (Archimandrite Gurias, Trans.) Beijing: Russian Mission.

The Summary of New Testament Salvation History: Life of Christ. (1861). (Archimandrite Gurias, Trans.) Beijing: Russian Mission.

Psalter. (1910). (Archimandrite Innocent, Trans.) Beijing: Russian Mission.

Saint Demetrius of Rostov. (1860). *Mirror of an Orthodox Confession.* (Archimandrite Gurias, Trans.) Beijing: Russian Mission.

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 Chinese Characters Used for Transliterating Slavonic*
2. Requester's name: *Yuri Shardt, Mitrophan Chin, Aleksandr Andreev*
3. Requester type (Member body/Liaison/Individual contribution): *Individual Contribution*
4. Submission date: *January 18th, 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: *Chinese Ideographs*
2. Number of characters in proposal: *19*
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? *Mitrophan Chin*
   If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: http://en.glyphwiki.org/font/gw989033.ttf
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; publishers of service books*
        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? *unknown*
    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 philological studies*

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)? *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? *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?
        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: