Proposal to Encode Chinese Characters Used for Transcribing Slavonic

Yuri Shardt, Mitrophan Chin, Aleksandr Andreev, Deborah Andersen

Introduction

This document details the proposed addition of Chinese ideographs used for transliterating Slavonic syllables. The following documents are enclosed in this folder:

1) **IRG_Unicode_13009.A.pdf**: This document, which presents the relevant evidence for the proposed new characters, as well as the proposal summary sheet.
2) **IRG_Unicode_13009.B.xlsx**, which is an Excel file containing the U-Source ID, Status, Unicode Codepoint, Radical-Stroke Count, Virtual Kangxi Dictionary Count, IDS Source, and Final Stroke Count. A summary of this table is presented as Figure 1 in this document.
3) **IRG_Unicode_13009.ttf**, which contains the TrueType font for the proposed characters. The characters are located in the Private Use Area starting with U+E000 with the codepoint incrementing by one for each character. The order of characters is the same as the order in Table 1.
4) **IRG_Unicode_13009.zip**, which is a zip folder containing bitmaps of all the proposed characters. The filenames of the individual bitmaps correspond with the U-Source ID of the corresponding character.

This submission is sponsored by the Script Encoding Initiative at UC Berkeley, an organization with Class C liaison to SC2. This submission is *not* sponsored by the Unicode Technical Committee or the Unicode Consortium. SEI is using the "UTC-" prefix as a U-Source identifier as a convenience, so there will be a well-known, stable source. The characters in this submission are listed in UAX #45 with these "UTC-" identifiers. Please note that a "UTC-" number in UAX #45 does not mean that the character with that identifier is sponsored by the Unicode Technical Committee. It is simply a catalogue number.

The following changes have been made since the last submission:

1) The request for sorting the letters by Slavonic phonological ordering has been withdrawn.
2) Additional information regarding the way the characters were originally created has been added.
3) Information regarding the use of these characters in Japanese transliterations has been added.
4) Information regarding vertical and horizontal display of the characters has been added.

<u>Background</u>

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 and Japanese. In order to accomplish this task for Chinese, Archimandrite Gurias (*né* Gregory 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. Similiarly, the translator Michael Hakugoku[3] adopted Archimandrite Gurias's series of characters for transcribing these sounds into the Japanese script using slightly different phonetic values, for example, 􏰁利爾 which is used to represent the sound *ri* in Chinese, represents a final *l* in Japanese. Since the same kanji can represent multiple readings depending on context in the Japanese language, Archimandrite Gurias's transliteration scheme was supplemented and quickly replaced by katakana in providing Slavonic-derived readings of proper names under the direction of Bishop Nikolai (*né* Ivan Dimitrovich Kasatkin)[4], the head of the first Russian Mission in Japan.

---

[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.
[3] Michael Hakugoku Kiyoshi (白極潔) was born in Sendai. There is limited information about his life, but it is known that he was in the 4th year of the seminary programme in 1878. It is also known that he later became a translator after graduation and that he died young.
[4] He later become the Archbishop of Tokyo in 1907.

Table 1 shows a summary of the proposed 20 characters along with their standard Mincho (print) and Kaishu (brush stroke) forms, and names, while Figure 1 presents a screenshot of the corresponding Excel spreadsheet file that contains additional required information. Of the proposed characters, all but the abbreviation for Jesus and 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 Innocent introduced an abbreviation character for the names *Christ* and *Jesus*, which were then used in the translations (see for example, Figure 20 and Figure 21). Finally, it can be noted that the simplified *ri* form was also developed and used.

It can be noted that Slavonic transliteration characters are optimised for vertical reading, that is, the characters are combined so that the right-hand character represents the first (initial) sound and the left-hand character represents the rime (or remainder), for example, in ⿰英微, the 微 represents the initial /v/ sound, while 英 represents the rime /in/. This representation is true for all the proposed characters, except for the abbreviations for Jesus and Christ. For these two characters, the individual elements are arranged vertically to form a single character. This optimisation for vertical display raises the question of what if anything should be done when these characters are displayed in left-to-right, horizontal text. Based on an analysis of the texts, the authors propose that nothing special should be done for the characters when they are displayed horizontal, for the following reasons:

1) Since these characters are primarily of historical interest, they will be found in most cases displayed in a vertical environment. One-off citation examples would not change the situation, even in a left-to-right, horizontal context.

2) There are no examples of such characters being used in a left-to-right, horizontal context and so practically speaking, this issue is moot, as there would not be any historical precedent for treating them any differently. Furthermore, it can be noted that in right-to-left, horizontal contexts, the letters are written as shown (as this would be expected).

3) If these characters are to be used in left-to-right, horizontal context, then it would be best to leave them as they currently stand, since it could be conceivable to treat them as a single unit that has lost any connection to its origin. In such cases, as there are no extant examples of left-to-right, horizontal usage, the proper stacking behaviour is undefined.

U-Source ID	Radical-Stroke Count	Virtual Kangxi Dictionary Count	IDS	Source	First Stroke Count
UTC-01179	140.18	1071.201	⿱英微	UTCDoc L2/13-009 1	2
UTC-01180	128.12	970.091	⿱耶格	UTCDoc L2/13-009 2	2
UTC-01181	9.14	120.111	⿱伊格	UTCDoc L2/13-009 3	1
UTC-01182	128.9	968.391	⿱耶克	UTCDoc L2/13-009 4	2
UTC-01183	9.11	116.041	⿱伊克	UTCDoc L2/13-009 5	2
UTC-01184	64.19	464.251	⿱拉爾	UTCDoc L2/13-009 6	4
UTC-01185	138.17	1014.011	⿱郎爾	UTCDoc L2/13-009 7	4
UTC-01186	78.16	584.091	⿱列爾	UTCDoc L2/13-009 8	2
UTC-01187	75.23	564.231	⿱楞爾	UTCDoc L2/13-009 9	2
UTC-01188	173.19	1383.061	⿱雷爾	UTCDoc L2/13-009 10	2
UTC-01189	115.16	861.451	⿱利爾	UTCDoc L2/13-009 11	2
UTC-01190	115.7	854.201	⿱利尔	UTCDoc L2/13-009 12	2
UTC-01191	15.22	133.561	⿱凌爾	UTCDoc L2/13-009 13	1
UTC-01192	122.28	950.371	⿱羅爾	UTCDoc L2/13-009 14	3
UTC-01193	170.23	1363.191	⿱隆爾	UTCDoc L2/13-009 15	3
UTC-01194	195.18	1480.431	⿱魯爾	UTCDoc L2/13-009 16	2
UTC-01195	128.8	968.221	⿱耶合	UTCDoc L2/13-009 17	2
UTC-01196	9.10	114.051	⿱伊合	UTCDoc L2/13-009 18	1
UTC-01197	1.6	78.131	⿳合一	UTCDoc L2/13-009 19	3
UTC-01198	1.8	78.171	⿳人伊一	UTCDoc L2/13-009 20	3

Figure 1: Summary of the Excel file containing the character information

Table 1: Summary of proposed new characters

Name	Composition	U-Source ID	Form Kaishu	Mincho	Comments
CJK SLAVONIC IDEOGRAPH VIN	□英微 82F1,5FAE	UTC-01179	英微	英微	equivalent to *vin* (вин) as in Навин (Nun, 那□英微)
CJK SLAVONIC IDEOGRAPH GE	□耶格 8036,683C	UTC-01180	耶格	耶格	equivalent to *ge* (ге) as in Нигер (Niger, 尼□耶格爾)
CJK SLAVONIC IDEOGRAPH GI	□伊格 4F0A,683C	UTC-01181	伊格	伊格	equivalent to *gi* (ги) as in Сергия (Sergius, 些爾□伊格乙)
CJK SLAVONIC IDEOGRAPH KE	□耶克 8036,514B	UTC-01182	耶克	耶克	equivalent to *ke* (ке) as in Кесария (Caesarea, □耶克薩□利爾亞)
CJK SLAVONIC IDEOGRAPH KI	□伊克 4F0A,514B	UTC-01183	伊克	伊克	equivalent to *ki* (ки) as in Езекия (Hezekiah, 耶捷□伊克亞)
CJK SLAVONIC IDEOGRAPH RA	□拉爾 62C9,723E	UTC-01184	拉爾	拉爾	equivalent to *ra* (ра) as in Израиль (Israel 伊斯□拉爾伊利)

Name	Composition	U-Source ID	Form Kaishu	Mincho	Comments
CJK SLAVONIC IDEOGRAPH RAN	□郎爾 90CE,723E	UTC-01185	郎爾	郎爾	equivalent to *ran* (ран) as in Аран (Haran, 哈爾□郎爾)
CJK SLAVONIC IDEOGRAPH RE	□列爾 5217,723E	UTC-01186	殑爾	殑爾	equivalent to *re* (ре) as in Назарет (Nazareth, 那匝□列爾特)
CJK SLAVONIC IDEOGRAPH REN	□楞爾 695E,723E	UTC-01187	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	UTC-01188	雷爾	雷爾	equivalent to *reia* (рея) as in назареянин (Nazarene, 那作□雷爾)
CJK SLAVONIC IDEOGRAPH RI	□利爾 5229, 723E	UTC-01189	稱爾	稱爾	equivalent to *ri* (ри) as in Христос (Christ, 合□利爾斯托斯); can also represent a final *l* in Japanese transcriptions.
CJK SLAVONIC IDEOGRAPH SIMPLIFIED RI	□利尔 5229,5C14	UTC-01190	称	称	Simplified variant of □利爾. Same usage as the traditional *ri*.

Name	Composition	U-Source ID	Form		Comments
			Kaishu	Mincho	
CJK SLAVONIC IDEOGRAPH RIN	☐凌爾 51CC,723E	UTC-01191	瀷	瀷	equivalent to *rin* (рин) as in Коринф (Corinth, 适☐凌爾福)
CJK SLAVONIC IDEOGRAPH RO	☐羅爾 7F85,723E	UTC-01192	羅爾	羅爾	equivalent to *ro* (ро) as in романский (Roman, ☐羅爾瑪)
CJK SLAVONIC IDEOGRAPH RON	☐隆爾 9686,723E	UTC-01193	隆爾	隆爾	equivalent to *ron* (рон) as in Аарон (Aaron, 阿阿☐隆爾)
CJK SLAVONIC IDEOGRAPH RU	☐魯爾 9B6F,723E	UTC-01194	鱺	鱺	equivalent to *ru* (ру) as in Иерусалим (Jerusalem, 耶☐魯爾薩利木)
CJK SLAVONIC IDEOGRAPH KHE	☐耶合 8036,5408	UTC-01195	聈	聈	equivalent to *khe* (хе) as in Сихем (Shechem, 西☐耶合木)
CJK SLAVONIC IDEOGRAPH KHI	☐伊合 4F0A,5408	UTC-01196	佮	佮	equivalent to *khi* (хи) as in Мелхий (Melchi, 芈利☐伊合乙)

| Name	Composition	U-Source ID	Form Kaishu	Mincho	Comments
CJK SLAVONIC IDEOGRAPH CHRIST	⿳合一 5408,4E00	UTC-01197	合	合	abbreviation for the word *Christ* (Христос, Khristos). The full form in Classical Chinese would be 合⿱利爾斯托斯.
CJK SLAVONIC IDEOGRAPH JESUS	⿱人伊一 4EBA,4F0A, 4E00	UTC-00198	侴	侴	abbreviation for the word *Jesus* (Иисус, Iisus). The full form in Classical Chinese would be 伊伊穌斯.
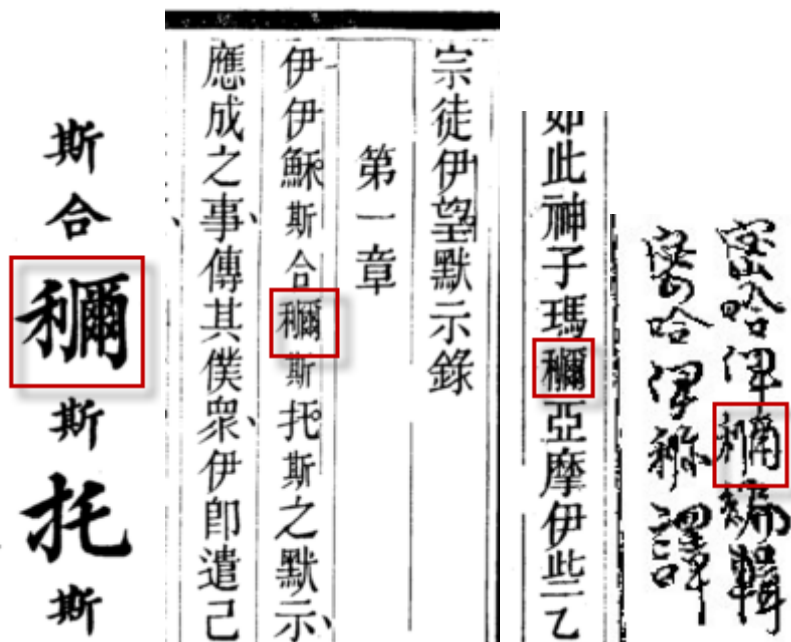
Figure 2: Examples of CJK SLAVONIC IDEOGRAPH RI (first, from left) from the cover page of the New Testament translated by Archimandrite Gurias (1864), (second) Book of Revelation showing use of the same character (1864), (third) from the front matter from the New Testament (1864), and (fourth) example from the cover page of a Japanese translation of the commentaries on the Gospel of St. Matthew (Archimandrite Michael of Kursk, Ming 15/16). It should be noted that in the last example the proposed character is used to represent an original final *l* in Michael.

Figure 3: Example of CJK SLAVONIC IDEOGRAPH SIMPLIFIED RI from (left) the front matter from the New Testament (1864) in chopmark seal style and (right) example from the cover page of a Japanese translation of the commentaries on the Gospel of St. Matthew (Archimandrite Michael of Kursk, Ming 15/16). It should be noted that in the last example the proposed character is used to represent an original final *l* in Michael.
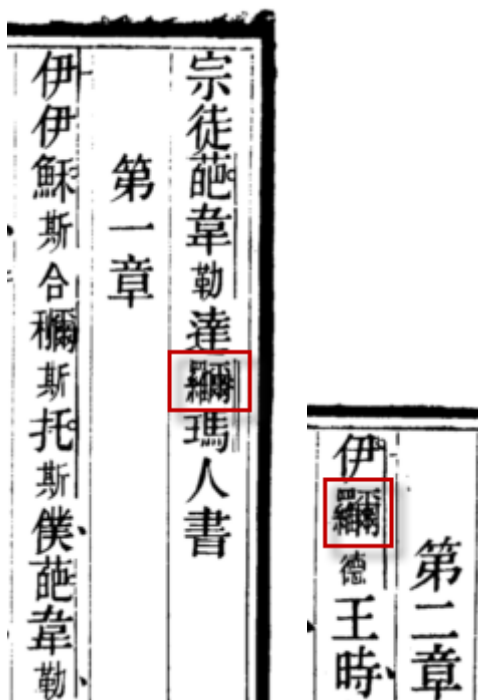
Figure 4: Example of CJK SLAVONIC IDEOGRAPH RO (left) from the start of the Epistle to the Romans (1864) and the name Herod at Matthew 2:1 from the New Testament (1864).
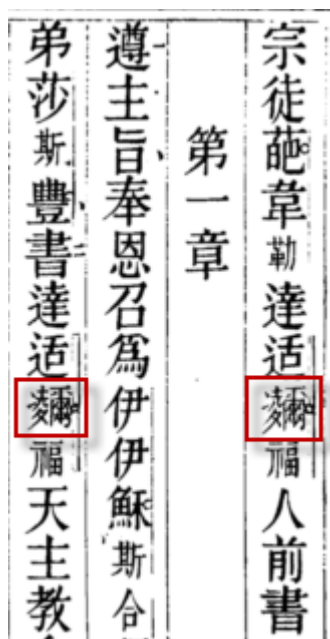

Figure 5: Example of CJK SLAVONIC IDEOGRAPH RIN from the start of the First Epistle to the Corinthians (1864).
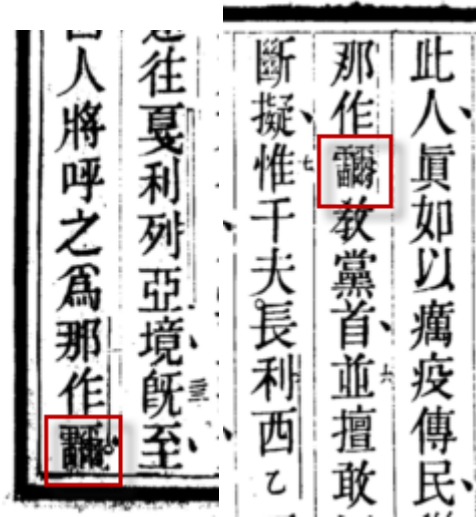
Figure 6: Examples of CJK SLAVONIC IDEOGRAPH REIA from (right) Matthew 2:23 and (left) Acts 24:5 (1864).
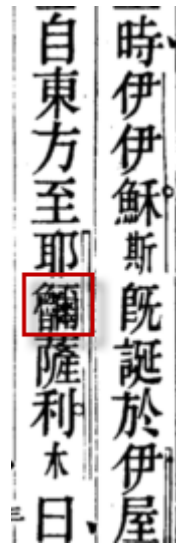

Figure 7: Example of CJK SLAVONIC IDEOGRAPH RU from Matthew 2:1 (1864).

Figure 8: Example of CJK SLAVONIC IDEOGRAPH RE from Matthew 2:23 (1864).

Figure 9: Examples of CJK SLAVONIC IDEOGRAPH RA (left) Romans 11:1 and (right) Matthew 1:8 (1864).

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

Figure 11: Example of CJK SLAVONIC IDEOGRAPH KHI from Matthew 2:18 (1864).

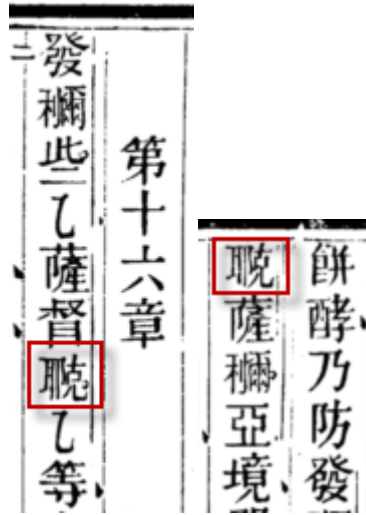Figure 12: Example of CJK SLAVONIC IDEOGRAPH GE from Mark 3:17 (1864).

Figure 13: Examples of CJK SLAVONIC IDEOGRAPH KE from (left) Matthew 16:1 and (right) Matthew 16:13 (1864).
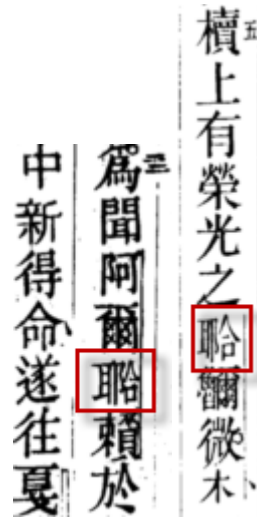

Figure 14: Examples of CJK SLAVONIC IDEOGRAPH KHE from (left) Matthew 2:22 and (right) Hebrews 9:5 (1864).
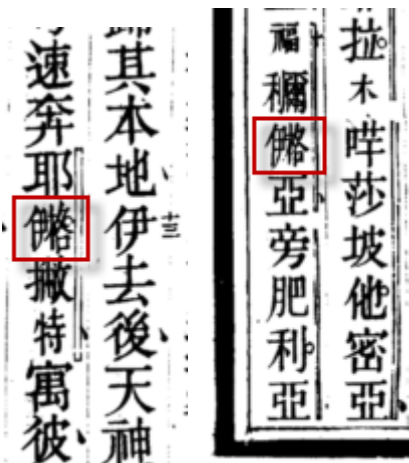
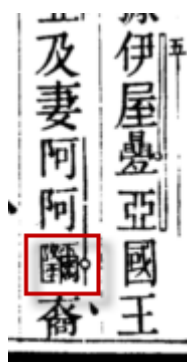Figure 15: Examples of CJK SLAVONIC IDEOGRAPH GI from (left) Matthew 2:13 and (right) Acts 2:10 (1864).

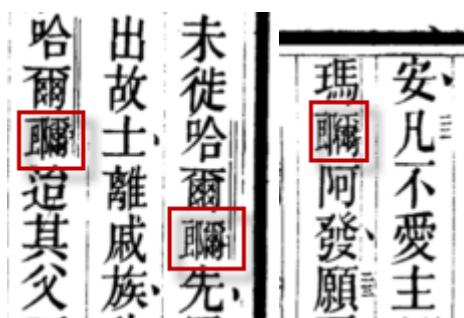Figure 16: Example of CJK SLAVONIC IDEOGRAPH RON from Luke 1:5 (1864).

Figure 17: Examples of CJK SLAVONIC IDEOGRAPH RAN from (left) Acts 7:2, 4 and (right) 1 Corinthians 16:22 (1864).

Figure 18: Example of CJK SLAVONIC IDEOGRAPH VIN from Hebrews 4:8 (1864).

Figure 19: Examples of CJK SLAVONIC IDEOGRAPH REN from (left) *The Summary of New Testament Salvation History: Life of Christ* translated by 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).
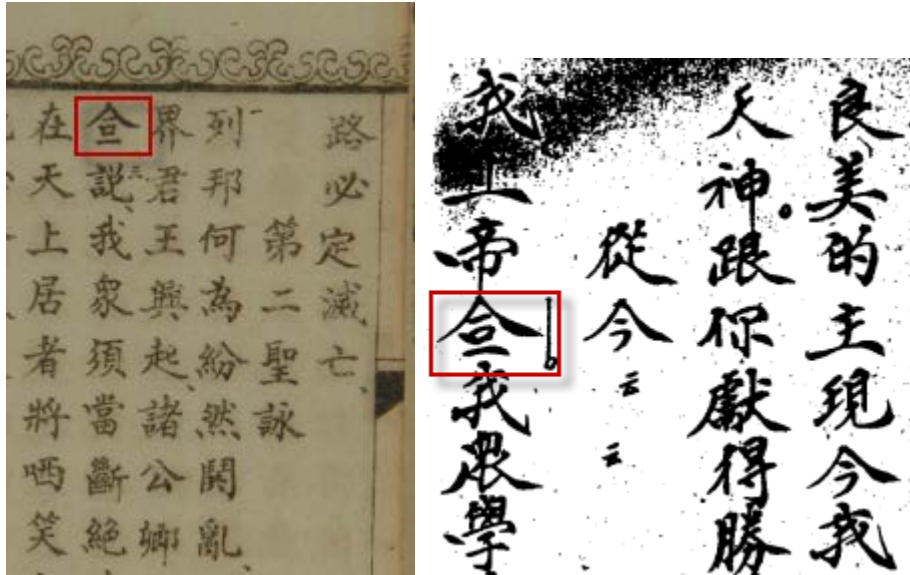
Figure 20: Examples of CJK SLAVONIC IDEOGRAPH CHRIST from (left) Psalm 2:2 from Bishop Innocent's translation of the Psalter into Mandarin (Psalter in Mandarin, 1910) and (right) extract from a manuscript of the Horologion translated into the vernacular Chinese language found in Harbin, China from 1930/40.
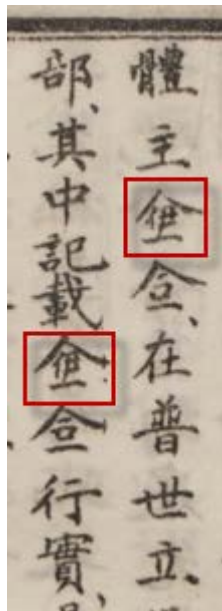

Figure 21: Example of CJK SLAVONIC IDEOGRAPH JESUS from Bishop Innocent's translation of the commentary on the Gospel of St. Matthew (1911).

Acknowledgement

References

Archimandrite Michael of Kursk. (Ming 15/16). *Commentaries on the Gospel of St. Matthew.* (M. Hakugoku, Trans.) Japan. Retrieved from http://kindai.ndl.go.jp/info:ndljp/pid/825673?itemId=info%3Andljp%2Fpid%2F825673&__lang=en

Commentary on the Gospel of St. Matthew. (1911). (Bishop Innocent, Trans.) Beijing: Russian Mission. Retrieved from http://orthodox.cn/bible/1911nt/Book_01/1911matthew_sm.pdf

New Testament in Chinese. (1864). (Archimandrite Gurias, Trans.) Beijing: Russian Mission. Retrieved from http://archive.wul.waseda.ac.jp/kosho/bunko08/bunko08_d0417/

Psalter in Mandarin. (1910). (Bishop Innocent, Trans.) Beijing: Russian Mission. Retrieved from http://orthodox.cn/bible/psalter/1910/index.html

Saint Demetrius of Rostov. (1860). *Mirror of an Orthodox Confession.* (Archimandrite Gurias, Trans.) Beijing: Russian Mission. Retrieved from http://catalogue.nla.gov.au/Record/1828830

The Summary of New Testament Salvation History: Life of Christ. (1861). (Archimandrite Gurias, Trans.) Beijing: Russian Mission. Retrieved from http://nla.gov.au/nla.gen-vn1815414

Annex F: IRG Repertoire Submission Summary Form

<table>
<tr><td colspan="2" align="center">

ISO/IEC JTC 1/SC 2/WG 2/IRG
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS OF CJK UNIFIED IDEOGRAPHS TO THE REPERTOIRE OF ISO/IEC 10646
Please fill in all the sections below.
Please read Principles and Procedures Document (P & P) from http://appsrv.cse.cuhk.edu.hk/~irg/irg31/IRGN1562.pdf
for guidelines and details before filling in this form.
Please ensure you are using the latest Form from http://appsrv.cse.cuhk.edu.hk/~irg/SubmissionForm.pdf .
See also http://appsrv.cse.cuhk.edu.hk/~irg/UCV.html **for latest *Unifiable Calligraphic Variations*.**

</td></tr>
</table>

A. Administrative

1. **IRG Project Code:**	SEI Urgently Needed Characters
2. **Title:**	Proposal to Encode Chinese Characters Used for Transcribing Slavonic
3. Requester's region/country name:	Script Encoding Initiative, UC Berkeley
4. Requester type (National Body/Individual contribution):	Liaison contribution (class C to SC2)
5. Submission date:	November 2013
6. Requested Ideograph Type (Unified or Compatibility Ideographs)	Unified Ideographs
If Compatibility, does requester have the intention to register them as IVS (See UTS #37) with the IRG's approval? (Registration fee will not be charged if authorized by the IRG.)	N/A
7. Request Type (Normal Request or Urgently Needed)	Urgently Needed
8. Choose one of the following:	
This is a complete proposal:	Yes
(or) More information will be provided later:	

B. Technical – General

1. Number of ideographs in the proposal:	*20*
2. Glyph format of the proposed ideographs: (128x128 "bmp" files or TrueType font file)	
If 'bmp' files, their file names are the same as their Source IDs?	
If TrueType font, all proposed glyphs are put into BMP PUA area?	*Yes*
If TrueType font, data for Source IDs vs. character codes are provided?	*Yes*
3. Source IDs:	
Do all the proposed ideographs have a unique, proper Source ID (country/region code and less than 9 alphanumeric characters)?	Yes
4. Evidence:	
a. Do all the proposed ideographs have the separate evidence document which contains at least one scanned image of printed materials (preferably dictionaries)?	Yes
b. Do all the printed materials used for evidence provide enough information to track them by a third party (ISBN numbers, etc.)?	Yes
5. Attribute Data Format: (Excel file or CSV)	*Excel File*

C. Technical - Checklist

Understandings of the Unification Checklist	
1. Has the requester read ISO/IEC 10646 Annex S and did the requester understand the unification policy?	*Yes*
2. Has the requester read the "Unifiable Calligraphic Variations" (contact IRG technical editor through the Rapporteurfor the latest one) and did the requester understand the unifiable variation examples?	*Yes*
3. Has the requester read this P&P document and did the requester understand the 5% rule?	*Yes*
Character-Glyph Duplication Checklist(http://www.itscj.ipsj.or.jp/sc2/open/pow.htm contains all the published ones and those under ballot**)**	
4. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the unified or compatibility ideographs of ISO/IEC 10646?	*Yes*
If yes, which version of ISO/IEC 10646 did requester check? (e.g. 10646:2003)	10646:2013
5. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the ideographs in Amendments of current ISO/IEC 10646? (As of 2009, Amendment 1, 4, 5, 6and 8 have CJK ideographs.)	*N/A*
If yes, which amendments did requester check?	N/A
6. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the ideographs in the current IRG working sets or proposed amendments of ISO/IEC 10646? (As of 2009, PDAM 6 and PDAM 8 have CJK ideographs.)	*Yes*
If yes, which draft amendments did requester check?	Extension F
7. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the ideographs in the current working M-set and D-set of the IRG? (Contact IRG chief editor and technical editor through the IRG Rapporteur for the newest list)	*Yes*
If yes, which document did requester check?	IRGN1921 (CJK Ext. F. v1.0)
8. Has the requester checked that any of the proposed ideographs is ***not unifiable*** with the over-unified or mis-unified ideographs in ISO/IEC 10646? (Check Annex E of this document).	*Yes*
9. Has the requester checked that any of the proposed ideographs ***has similar ideograph(s)*** with the ideographs in the current standardized or working set mentioned above?	*Yes*
10. Has the requester checked that any of the proposed ideographs ***has variant ideograph(s)*** with the ideographs in the current standardized or working set mentioned above?	*Yes*
Attribute Data Checklist	
11. Do all the proposed ideographs have attribute data such as the KangXi radical code, stroke count and first stroke?	*Yes*
12. Are there any simplified ideographs (ideographs that are based on the policy described in 簡化字總表) in the proposed ideographs?	*No*
If YES, does your proposal include proper simplified/traditional indication flag for each proposed ideograph in attribute data?	
13. Do all the proposed ideographs have the document page number of evidence documents in attribute data?	*Yes*
14. Do all the proposed ideographs have the proper Ideographic Description Sequence (IDS) in attribute data?	*Yes*
If NO, how many proposed ideographs do not have the IDS?	
15. If the answer to question 9 or 10 is yes, do the attribute data include any information on similar/variant ideographs for the proposed ideographs?	*Yes*