

Title: Proposal to Encode 20 Additional Kanbun Marks

Author: Wang Yifan

Date: 2020-01-06

## 1. Introduction

This document augments L2/20-232 (Wang, 2020) with various marginal Kanbun marks found in early modern literature, as well as fuller data and information of those newly proposed marks, which would be needed upon inclusion to Unicode. A revised Proposal Summary Form is attached at the end.

## 2. Addition from Early Modern Annotation

### 2.1. *Ōbun Kundoku*

*Ōbun kundoku* 欧文訓読 was a once popular method practiced from the 18<sup>th</sup> through the 19<sup>th</sup> century, that reads European languages' text in word-by-word transposed Japanese translation, applying already established *kanbun kundoku* rules with various modifications. Although the contact between Japanese and Europeans started in the 16<sup>th</sup> century, the communication was shut down by the isolationist policy in the next century. Only after 1720, when the importation of Dutch books is permitted by the shogunate, the interest and studies toward the Dutch language gather momentum. Since then, the practice had been found handy especially for novice learners and applied to multiple European languages after the abandonment of isolation. Morioka (1999, p. 127) cites the last *ōbun kundoku* style textbook published in as late as 1906.

### 2.2. New Marks and Sources

This document lists 20 marks in total, of which 9 are already introduced in L2/20-232. We will provide detailed information for the remaining 11 characters in this section.

- Addition in Level 1 marks
  - 十 (tenth)
- Addition in Level 3 marks
  - 庚・辛・壬・癸・子・丑・寅・卯・辰 (seventh through fifteenth)
- Supplementary
  - 廿 (twenty in serial numbering)

These are attested in following documents:

- a) *Oranda yakusen* 和蘭譯筌 (1785) by Maeno Ryōtaku 前野良沢 [Fig. 1, Fig. 2]
- b) *Pineo-shi genpan eibunten chokuyaku* ピ子ヲ氏原板英文典直譯 (1870), translated and annotated by Nagashima Teijirō 永嶋貞次郎 [Fig. 3]

### 2.3. Detailed Information

- (1) The seventh example of the chapter “Rankwatei Yakubunshiki” 蘭化亭譯文式 of *Oranda Yakusen* (Fig. 1) contains new annotation marks 庚 to 辰, as transcribed on the top right of the page (Numata et al., 1976; p. 125). Even though never printed, this book is one of the most influential handbooks in the early stage of Dutch language studies in Japan. The peculiar methodology demonstrated in the book is, rather than annotating Dutch words directly, assigning a Chinese phrase to a Dutch word and reading Chinese phrases accordingly in Japanese. To compensate for the grammatical and graphical scattering caused by the double-deck translation (the European text goes left-to-right, top-to-bottom while East Asian predominantly top-to-bottom, right-to-left), the author relinquishes usual stratified marking of *kanbun kundoku* and simply links each chunk of phrase (not always equal a Dutch word) in a sequential order<sup>1</sup>. That is the reason why such higher level 3 marks appear in this text: he uses level 1 marks (一, 二, 三...) in some shorter examples, but the series cannot handle more than ten marks; level 3 represents ten heavenly stems 十干 which can be extended with another twelve earthly branches 十二支 if needed, consisting of totally 22 members. This expansion is theoretically acknowledged in *Kanbun* (Kotajima, 2009; p. 34), just too rare to appear in the mainstream *kanbun kundoku*.

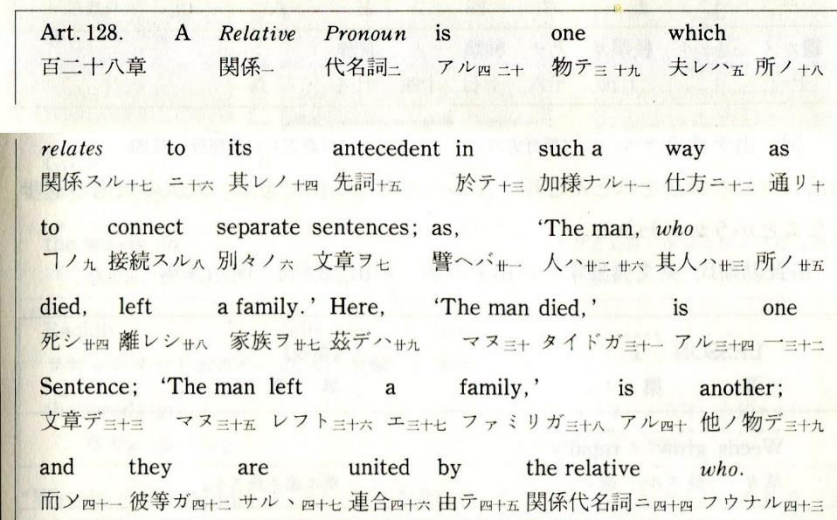
As our first evidence technically lacks 卯 in any witness we have found (should have been under “Isaak van der Putte”), as

well as 辛 in our manuscript image, we supplement them with the second example of the same chapter (Fig. 2; transcription on the middle right (Numata et al., 1976; p. 122)).

- (2) The usage of 十 and 廿 marks is taken from *Pineo-shi genpan eibunten chokuyaku* (Fig. 3), which provides Japanese gloss and reading order corresponding to almost every single word written in the main body of Timothy Stone Pinneo’s *Primary Grammar of the English Language*. The book as such does not include English text and is meant to be read with the original English textbook (Fig. 4) side by side. The parallelized transcription with English and Japanese is given below from



<sup>1</sup> This is not a standard practice of *kanbun kundoku*, even at that time. He also has works written in Classical Chinese, which are annotated in a way hardly different with that of today (Oita Prefecture Ancient Sages Historical Archives, 2009).



Morioka (1999, pp. 112–113).

The latest stage of *ōbun kundoku* utilizes much more straightforward serial numbering<sup>2</sup> to indicate the reading order in each sentence (words that need to be parsed twice in Japanese are marked twice). It thus eliminates the need to devise a special mechanism to cope with European grammatical features like relatives, syllepses, postmodifying elements and such, while allows for indeterminate number of words per sentence.

As a result, however, the system stands on the borderline between Kanbun marking and plain numbering in Chinese numerals. Though it is built upon (mostly) a subset of Kanbun marks, its semantics differs from traditional Kanbun in the point that each single mark is not self-standing but functioning as part of number spelling. Especially, 廿 is heterodox in terms of traditional Kanbun rules because it would have to go through “eleven” to “nineteen” before reaching it, while they have no single-character representation in Chinese characters<sup>3</sup>.

### 3. Proposal

#### 3.1. All Glyphs and Block(s)

Along with L2/20–232, two columns of 16 code points will be needed to contain the new characters. The suggested arrangement and representative glyphs are as follows, with tentative code points:

Code Point	Glyph	Name
U+XXXX0	五	IDEOGRAPHIC ANNOTATION FIVE MARK

<sup>2</sup> Actually, mid stage *ōbun kundoku* generally employed much more moderate annotation strategies that involve less higher-number marks seen in this proposal, but instead introduced some eccentric marking method such as intersecting jump, double anchoring, or single word parsed five times to process various grammatical complication, most visibly in *Kunten oranda bunten* (1857).

<sup>3</sup> Note that few textbooks explicitly mention the upper bound of each series, and some books, e.g. Kotajima and Yuki (2011) states that level 1 is an open set, which may entail that compound numerals are acceptable.

U+XXXX1	六	IDEOGRAPHIC ANNOTATION SIX MARK
U+XXXX2	七	IDEOGRAPHIC ANNOTATION SEVEN MARK
U+XXXX3	八	IDEOGRAPHIC ANNOTATION EIGHT MARK
U+XXXX4	九	IDEOGRAPHIC ANNOTATION NINE MARK
U+XXXX5	十	IDEOGRAPHIC ANNOTATION TEN MARK
U+XXXX6	乾	IDEOGRAPHIC ANNOTATION CREATIVE HEAVEN MARK
U+XXXX7	坤	IDEOGRAPHIC ANNOTATION RECEPTIVE EARTH MARK
U+XXXX8	戊	IDEOGRAPHIC ANNOTATION FIFTH MARK
U+XXXX9	己	IDEOGRAPHIC ANNOTATION SIXTH MARK
U+XXXXA	庚	IDEOGRAPHIC ANNOTATION SEVENTH MARK
U+XXXXB	辛	IDEOGRAPHIC ANNOTATION EIGHTH MARK
U+XXXXC	壬	IDEOGRAPHIC ANNOTATION NINTH MARK
U+XXXXD	癸	IDEOGRAPHIC ANNOTATION TENTH MARK
U+XXXXE	子	IDEOGRAPHIC ANNOTATION ELEVENTH MARK
U+XXXXF	丑	IDEOGRAPHIC ANNOTATION TWELFTH MARK
U+YYYY0	寅	IDEOGRAPHIC ANNOTATION THIRTEENTH MARK
U+YYYY1	卯	IDEOGRAPHIC ANNOTATION FOURTEENTH MARK
U+YYYY2	辰	IDEOGRAPHIC ANNOTATION FIFTEENTH MARK

U+YYYY3		(reserved; for possible SIXTEENTH MARK to come)
U+YYYY4		(reserved; for possible SEVENTEENTH MARK to come)
U+YYYY5		(reserved; for possible EIGHTEENTH MARK to come)
U+YYYY6		(reserved; for possible NINETEENTH MARK to come)
U+YYYY7		(reserved; for possible TWENTIETH MARK to come)
U+YYYY8		(reserved; for possible TWENTY-FIRST MARK to come)
U+YYYY9		(reserved; for possible TWENTY-SECOND MARK to come)
U+YYYYA		(currently unused)
U+YYYYB		(currently unused)
U+YYYYC		(currently unused)
U+YYYYD		(currently unused)
U+YYYYE		(currently unused)
U+YYYYF	廿	IDEOGRAPHIC ANNOTATION TWENTY MARK

We notice a suggestion regarding the block placement, L2/20-276 (Moist, 2020), that it take up the column of U+2FE0 to U+2FEF in BMP, “allowing the block to ‘fit neatly’ within the plane while situating it with similar blocks”. It may be reasonable in consideration of FIVE MARK 五 and FIFTH MARK 戊, which are among official repertoire of pedagogical Kanbun marks (see L2/20-232) and hence have been continually appeared in various contemporary textbooks to be an active modern usage. Meanwhile, we do not have strong preference by which plane they are accommodated, mostly because the inventory has exceeded the one-column frame. Either way, the first 16 code points can be separated from the rest if necessary. **XXXX** and **YYYY** in the table stand for tentative high digits of code points.

We believe that most of those characters bear as much usefulness as existing Kanbun marks (see the discussion in L2/20-232). The glyphs will look as if placed on the top left corner of an em square, consistent with existing ones.

### 3.2. UnicodeData.txt Addition

```
XXXX0;IDEOGRAPHIC ANNOTATION FIVE MARK;No;0;L;<super> 4E94;;;5;N;;;;;
XXXX1;IDEOGRAPHIC ANNOTATION SIX MARK;No;0;L;<super> 516D;;;6;N;;;;;
XXXX2;IDEOGRAPHIC ANNOTATION SEVEN MARK;No;0;L;<super> 4E03;;;7;N;;;;;
XXXX3;IDEOGRAPHIC ANNOTATION EIGHT MARK;No;0;L;<super> 516B;;;8;N;;;;;
XXXX4;IDEOGRAPHIC ANNOTATION NINE MARK;No;0;L;<super> 4E5D;;;9;N;;;;;
XXXX5;IDEOGRAPHIC ANNOTATION TEN MARK;No;0;L;<super> 5341;;;10;N;;;;;
XXXX6;IDEOGRAPHIC ANNOTATION CREATIVE HEAVEN MARK;So;0;L;<super> 4E7E;;;N;;;;;
XXXX7;IDEOGRAPHIC ANNOTATION RECEPTIVE EARTH MARK;So;0;L;<super> 5764;;;N;;;;;
XXXX8;IDEOGRAPHIC ANNOTATION FIFTH MARK;So;0;L;<super> 620A;;;N;;;;;
XXXX9;IDEOGRAPHIC ANNOTATION SIXTH MARK;So;0;L;<super> 5DF1;;;N;;;;;
XXXXA;IDEOGRAPHIC ANNOTATION SEVENTH MARK;So;0;L;<super> 5E9A;;;N;;;;;
XXXXB;IDEOGRAPHIC ANNOTATION EIGHTH MARK;So;0;L;<super> 8F9B;;;N;;;;;
XXXXC;IDEOGRAPHIC ANNOTATION NINTH MARK;So;0;L;<super> 58EC;;;N;;;;;
XXXXD;IDEOGRAPHIC ANNOTATION TENTH MARK;So;0;L;<super> 7678;;;N;;;;;
XXXXE;IDEOGRAPHIC ANNOTATION ELEVENTH MARK;So;0;L;<super> 5B50;;;N;;;;;
XXXXF;IDEOGRAPHIC ANNOTATION TWELFTH MARK;So;0;L;<super> 4E11;;;N;;;;;
YYYY0;IDEOGRAPHIC ANNOTATION THIRTEENTH MARK;So;0;L;<super> 5BC5;;;N;;;;;
YYYY1;IDEOGRAPHIC ANNOTATION FOURTEENTH MARK;So;0;L;<super> 536F;;;N;;;;;
YYYY2;IDEOGRAPHIC ANNOTATION FIFTEENTH MARK;So;0;L;<super> 8FB0;;;N;;;;;
YYYYF;IDEOGRAPHIC ANNOTATION TWENTY MARK;So;0;L;<super> 5EFF;;;20;N;;;;;
```

While they can have compatibility decomposition `<super>` to agree with existing Kanbun block characters, the well-practiced convention in horizontal environment is subscripting them (okurigana is usually superscripted)<sup>4</sup>, which would make `<sub>` a better alternative to reflect the actual usage.

---

<sup>4</sup> For example, see <http://id.nii.ac.jp/1579/00000313/>, <https://ja.wikibooks.org/wiki/高等学校古文/漢文の読み方>, <https://lscholar.hatenablog.com/entry/2015/04/21/132723> etc.

### 3.3. Scripts.txt Addition

The Script property should be **Common** for all proposed characters, which would look like:

```
XXXX0..XXXX5 ; Common # No
XXXX6..XXXXF ; Common # So
YYYY0..YYYY2 ; Common # So
YYYYF          ; Common # So
```

### 3.4. ScriptExtensions.txt Addition

The Script\_Extensions property should be **Hani** for all proposed characters, which would look like:

```
XXXX0..XXXX5 ; Hani # No
XXXX6..XXXXF ; Hani # So
YYYY0..YYYY2 ; Hani # So
YYYYF          ; Hani # So
```

### 3.5. EastAsianWidth.txt Addition

The East\_Asian\_Width property should be **W** for all proposed characters, which would look like:

```
XXXX0..XXXX5;W # No
XXXX6..XXXXF;W # So
YYYY0..YYYY2;W # So
YYYYF;W         # So
```

## Acknowledgements

On top of what is already mentioned in L2/20-232, we express our gratitude to Dr. Kiyonori Nagasaki and Prof. Masahiro Shimoda for the SAT text data and their continuous backup, Dr. Ken Lunde for his generous help and resources including the courtesy to let me use his font, and UTC members for the valuable advice.

## References

- Kotajima, Yosuke 古田島洋介 (2009). *Kore nara wakarū kaeriten: nyūmon kara ōyō made* これならわかる返り点: 入門から応用まで. Tokyo: Shintensha.
- Kotajima, Yosuke 古田島洋介 and Yoshinobu Yuki 湯城吉信 (2011). *Kanbun kundoku nyūmon* 漢文訓読入門. Tokyo: Meiji Shoin.
- Kunten oranda bunten* 訓點和蘭文典 (1857). Sōsetsukwan 総攝館.  
[https://www.wul.waseda.ac.jp/kotenseki/html/bunko08/bunko08\\_c0543/](https://www.wul.waseda.ac.jp/kotenseki/html/bunko08/bunko08_c0543/).
- Maeno, Ryōtaku 前野良澤 (n.d.). *Oranda yakusen* 和蘭譯筌. Manuscript. <https://rmda.kulib.kyoto-u.ac.jp/item/rb00001412>.
- Moist, Mika (2020). “Suggestion concerning tentative roadmap placement of Kanbun Extended block” [L2/20-276]. <https://www.unicode.org/L2/L2020/20276-kanbun-roadmap.pdf>.

- Morioka, Kenji (1999). *Ōbun kundoku no kenkyū: ōbunmyaku no keisei* 欧文訓読の研究: 欧文脈の形成. Tokyo: Meiji Shoin.
- Numata, Jirō 沼田次郎, Akira Matsumura 松村明 and Shosuke Sato 佐藤昌介 (1976). *Yōgaku I* 洋学 上. In: *Nihon shisō taikai* 日本思想大系 series. Tokyo: Iwanami Shoten.
- Oita Prefecture Ancient Sages Historical Archives 大分県立先哲史料館 (2009). *Maeno Ryōtaku shiryōshū* 前野良沢資料集 (vol. 1). In: *Ōita-ken Sentetsu Sōsho* 大分県先哲叢書 series. Oita: Oita Prefectural Board of Education.
- Pinneo, T. S. (1854). *Pinneo's Primary Grammar of the English Language: For Beginners* (Revised and Enlarged). New York: Van Antwerp, Bragg & Co.  
<https://archive.org/details/pinprimgrammar00pinneo/>.
- Pinneo, T. S. (1870). *Pineo-shi genpan eibunten chokuyaku* ピ子ヲ氏原板英文典直譯 (Nagashima, Teijirō 永嶋貞次郎 Trans.; vol. 1). Tokyo: Shōkodō 尚古堂.  
[https://www.wul.waseda.ac.jp/kotenseki/html/bunko08/bunko08\\_c0707/](https://www.wul.waseda.ac.jp/kotenseki/html/bunko08/bunko08_c0707/).
- Wang, Yifan (2020). "Proposal to Encode 9 Additional Kanbun Marks" [L2/20-232].  
<https://www.unicode.org/L2/L2020/20232-kanbun-additions.pdf>.



二分

Deze laatste Druk 同下  
甲此一 未二新 丙刻三 丁  
 ichtig Verzien en van Hoitsee  
卷四 見 庚 者 誤八  
 Gezigt tot Amsterdam  
清之 夜九 五 名都蘭和  
 Bij Isaak van der Pitte  
從 發行 名 姓  
 Baer Verkooper  
書  
 M D C C X I  
今時 MDCCXI  
四千七百 子 五百四十 五 十 年

Fig. 1 Example 7 of "Rankwatei Yakubunshiki" in manuscript (Maeno, n.d.)

Leer bekwaam om alle レツ  
テ  
ル  
コ  
ニ  
ス  
ト  
題  
言  
中  
大寅 (空適) 子應得益 卯 全癸 諸甲

Leer peem in korten tyd op (右就成)  
子〇△ 〇 (短不及丑) 時

De gemakkelijkste wijze te 〇  
易簡 法丁

Leeren 用字 〇 及 讀書 〇  
學字 巳

Schrijven. 書字  
辛

Fig. 2 Example 2 of "Rankwatei Yakubunshiki" in manuscript (Maeno, n.d.)

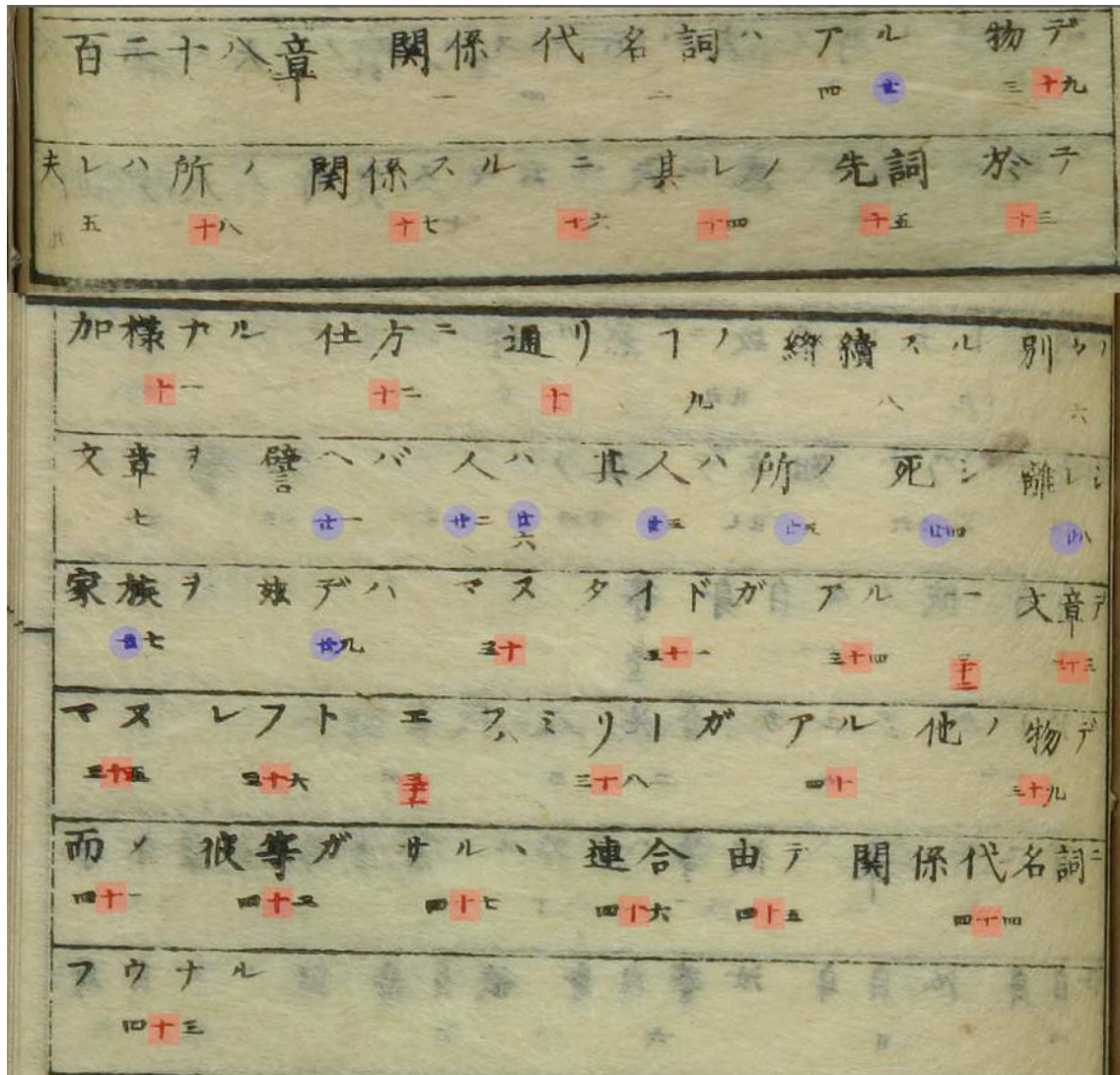


Fig. 3 Article 128, *Pineo-shi genpan eibun ten chokuyaku* (Pinneo, 1870; folio 37)

red squares highlight 十, blue circles 廿

ART. 128. A *Relative Pronoun* is one which *relates* to its antecedent in such a way as to connect separate sentences; as, 'The man, *who* died, left a family.' Here, 'The man died,' is one sentence; 'The man left a family,' is another; and they are united by the relative *who*. (See the author's *Analytical Grammar*.)

Fig. 4 Article 128, *Pinneo's Primary Grammar of the English Language* (Pinneo, 1854; p. 67)

**ISO/IEC JTC 1/SC 2/WG 2  
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS  
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646<sup>1</sup>.**

**Please fill all the sections A, B and C below.**

Please read Principles and Procedures Document (P & P) from <http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

**A. Administrative**

1. Title:	<i>Proposal to Encode 20 Additional Kanbun Marks</i>
2. Requester's name:	<i>Wang Yifan</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>2021-01-06</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<input checked="" type="checkbox"/>
(or) More information will be provided later:	<input type="checkbox"/>

**B. Technical – General**

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	<input type="checkbox"/>
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<input checked="" type="checkbox"/>
Name of the existing block:	<i>Kanbun</i>
2. Number of characters in proposal:	<i>20</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	
A-Contemporary	<input checked="" type="checkbox"/>
B.1-Specialized (small collection)	<input type="checkbox"/>
B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>
D-Attested extinct	<input type="checkbox"/>
E-Minor extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>
G-Obscure or questionable usage symbols	<input type="checkbox"/>
4. Is a repertoire including character names provided?	<input checked="" type="checkbox"/>
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?	<input checked="" type="checkbox"/>
b. Are the character shapes attached in a legible form suitable for review?	<input type="checkbox"/>
5. Fonts related:	
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?	<i>The requester or Dr. Ken Lunde</i>
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):	<i>747.neutron@gmail.com; https://github.com/unicode-org/cjk-symbols</i>
6. References:	
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?	<input checked="" type="checkbox"/>
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?	<input checked="" type="checkbox"/>
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)?	<input type="checkbox"/> <i>No</i>

**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.

<sup>1</sup> Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

**C. Technical - Justification**

1. Has this proposal for addition of character(s) been submitted before? If YES explain	<i>Expansion of L2/20-232</i>	Yes
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.)? If YES, with whom? If YES, available relevant documents:	<i>Japanese font technical experts</i>	Yes
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	<i>See proposal</i>	Yes
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>See proposal</i>	Educational
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	<i>See proposal</i>	Yes
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference:		No
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?		No
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference:		No
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference:		No
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? If YES, is a rationale for its inclusion provided? If YES, reference:	<i>See proposal and L2/20-232</i>	Yes Yes
11. Does the proposal include use of combining characters and/or use of composite sequences? 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:		No
12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary)		No
13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference:		No