Proposal to Encode Missing Japanese Kana

Abraham Gross
2020-01-05
abrahambgross@gmail.com

Introduction

The Japanese syllabary is composed of 102 characters collectively called "kana". Of those, 51 characters are in hiragana (平仮名・ひらがな) and 51 in katakana (片仮名・カタカナ). The current version of Unicode contains 3 extinct hiragana characters — 𛀁 (YE), ゐ (WI), and ゑ (we), and 2 extinct katakana characters — ヰ (WI), and ヱ (WE). However, there are still 2 hiragana characters — 𛀆 (YI) and 𛀊 (WU), and 3 katakana characters — 𛄠 (YI), 𛄝 (YE), and 𛄢 (WU), missing from Unicode. These characters were primarily created to reach the coveted 50 character mark of the "50-sound chart (五十音図)", and to represent these sounds in linguistic literature.

The は行 sounds in Japanese went through a lot of changes throughout the ages. It went through different changes depending on where in the word it was. In ancient times, all は行 sounds were pronounced /pV/ ("V" being any of the Japanese vowels). Then, during the Nara period, it shifted to /ɸV/ when it was in the middle or at the end of a word. Between the Heian and Kamakura periods, /ɸV/ changed to /βV/→/βV/→/V/ (except /βa/ which didn't lose the /β/). This means that at some point, words like「思う (which used to be おもふ)」and「扇 (which used to be あふぎ)」were pronounced「おもげ」/omoβu/ and「あけぎ」/aβugi/, demonstrating that the WU sound existed in Japanese.

Current Use

In colloquial Japanese, many diphthongs disappeared. So, words like でかい (DEKAI) and やばい (YABAI) became でけえ (DEKEE) and やべえ (YABEE), and words like わるい (WARUI) and さむい (SAMUI) became わりい (WARII) and さみい (SAMII). With these changes, words like はやい (HAYAI) became はⁿえ (HAYEE), and words like かゆい (KAYUI) became か𛀁い (KAYII). Having these characters available for writing would be invaluable as a way to represent these sounds in Japanese, for transcribing into Japanese, for digitizing old books, and for Japanese scholars.

Proposed Characters

Codepoint	Name	Character	Kanji Derivation
1B129	HIRAGANA LETTER YI	ｌ𛄝	以 (U+4EE5)
1B12A	HIRAGANA LETTER WU	𛄞	紆 (U+7D06)[1]
1B12C	KATAKANA LETTER YI	レ	以 (以 U+4EE5)
1B12D	KATAKANA LETTER YE	𛄲	衣 (衣 U+8863)[2]
1B12E	KATAKANA LETTER WU	于	宇 (宇 U+5B87)
1B163	KATAKANA LETTER SMALL WU	于	

Of the proposed characters, the following 4 are the most important (shown in both sans and serif fonts):

𛄞レ𛄲于
𛄞レ𛄲于

1　平仮名のわ行うの字源, 2019, https://kobunworld.blog.fc2.com/blog-entry-1.html

2　An alternative theory suggests that 𛄲 is a ligature of イ and エ. While plausible, it seems more probable to come from 衣 as shown in pages 7, 8, and 14.

1B12A;HIRAGANA LETTER WU;Lo;0;L;;;;;N;;;;;
1B12C;KATAKANA LETTER YI;Lo;0;L;;;;;N;;;;;
1B12D;KATAKANA LETTER YE;Lo;0;L;;;;;N;;;;;
1B12E;KATAKANA LETTER WU;Lo;0;L;;;;;N;;;;;

The following 2 characters have also been included as potential additions (in sans and serif fonts):

𛄩 𛅣
𛄩 𛅣

1B129;HIRAGANA LETTER YI;Lo;0;L;;;;;N;;;;;
1B163;KATAKANA LETTER SMALL WU;Lo;0;L;;;;;N;;;;;

These two characters are included for multiple reasons:

- ャ (small YA), ュ (small YU), ョ (small YO), and ヮ (small WA), can be attached to other kana in order to create a consonant cluster or to switch the sound (this is called 拗音・ようおん・*YOU'ON*). For example, ニャ is pronounced "NYA" not "NIYA", チュ is pronounced "CHU" not "CHIYU", and クヮ is pronounced "KWA "not "KUWA". With the addition of the character 𛅣 (small WU), it would finally be possible to write things such as クヰ "KWU" when transcribing into Japanese. (Evidence of use shown below.)

The proposed 𛄩 is a different character than the hentaigana 𛀆 U+1B006 HENTAIGANA LETTER I-1 due to it being the "neat" version, just like many other kana are the "neat" version of hentaigana. Examples:

- 久 U+1B02B HENTAIGANA LETTER KU-1　→ く U+304F HIRAGANA LETTER KU
- 計 U+1B036 HENTAIGANA LETTER KE-5　→ け U+3051 HIRAGANA LETTER KE
- 之 U+1B044 HENTAIGANA LETTER SI-1　→ し U+3057 HIRAGANA LETTER SI
- 八 U+1B09E HENTAIGANA LETTER HA-1　→ ハ U+30CF KATAKANA LETTER HA
- 津 U+1B06A HENTAIGANA LETTER TU-2　→ つ U+3064 HIRAGANA LETTER TU
- 留 U+1B0F9 HENTAIGANA LETTER RU-2　→ る U+308B HIRAGANA LETTER RU
- 惠 U+1B112 HENTAIGANA LETTER WE-1　→ ゑ U+3091 HIRAGANA LETTER WE

Just like these listed simplified characters, the proposed 𛄩 should be the neat version of the

character ⍵ (U+1B006 HENTAIGANA LETTER I-1). To elaborate, ⟨ ⟩[3], ⟨ ⟩[4], and ⟨ ⟩[5], are examples of the character ⍵ written neatly. As can be seen, the left-most line and the right arch are disconnected when written neatly, as well as the part shown in red here ⟨ ⟩ not being drawn because that is just the pen not lifting while going to the next character and its not actually part of the character.

If the consortium decides to not add ⍵, then I would like to suggest another option — an alias for U+1B006 HENTAIGANA LETTER I-1 to become HIRAGANA LETTER YI just like how U+1B001 HENTAIGANA LETTER E-1 became HIRAGANA LETTER ARCHAIC YE. (Proof of usage as "YI" rather than "I" is provided below.)

While 干 (WU) does look really similar to the Chinese character 于 (U+4E8E CJK UNIFIED IDEOGRAPH-4E8E), the similarity in shape is just as superficial as that of l (U+006C LATIN SMALL LETTER L), I (U+0049 LATIN CAPITAL LETTER I), and | (U+007C VERTICAL LINE).

The suggested names of the proposed characters were based on the names of other kana, and the character codepoint positions were chosen based on the positions of the proposed characters U+1B127 HIRAGANA LETTER SMALL KO, U+1B12B KATAKANA LETTER SMALL KO in the Kana Extended-A block, and also based on the positions of the characters in the Small Kana Extension block.

※ All of the character codepoints, names, and properties given are just suggestions.

---

3   日本文典. 巻之上, 1897, http://dl.ndl.go.jp/info:ndljp/pid/864063/6

4   単語独稽古, 1874, http://dl.ndl.go.jp/info:ndljp/pid/863316/2

5   仮名遣の栞,1892, http://dl.ndl.go.jp/info:ndljp/pid/862214/7

Sorting Order

Proposed Character	Should go after	And before
⿻ HIRAGANA LETTER YI	や U+3084 HIRAGANA LETTER YA	ゆ U+3085 HIRAGANA LETTER SMALL YU
け HIRAGANA LETTER WU	ゐ U+3090 HIRAGANA LETTER WI	ゑ U+1B151 HIRAGANA LETTER SMALL WE
⿻ KATAKANA LETTER YI	ヤ U+30E4 KATAKANA LETTER YA	ュ U+30E5 KATAKANA LETTER SMALL YU
イ KATAKANA LETTER YE	ユ U+30E6 KATAKANA LETTER YU	ョ U+30E7 KATAKANA LETTER SMALL YO
于 KATAKANA LETTER SMALL WU	ヰ U+30F0 KATAKANA LETTER WI	于 KATAKANA LETTER WU
于 KATAKANA LETTER WU	于 KATAKANA LETTER SMALL WU	ヱ U+1B165 KATAKANA LETTER SMALL WE

So in the end, to match the convention, the sorting order should be :
ゃやⱴゅゆⱴょよゎわゐゑゑけゑをを
ャヤⱴュユイョヨヮワヴゎヰギ于于ェヱゴョヲヺ

Evidence of Use

中と下とに在てふに紛るゝもの		
平假名	片假名	漢字
いもうと	イモ子ト	妹
けういく	ケ子イク	教育

平假名	片假名	漢字
たうべ	タ子ベ	賜給
さうま	サ子マ	相馬
玄うきく	シ子キク	蹴鞠
かうし	カ子シ	格子
かうつけ	カ子ツケ	上野
かうばし	カ子バシ	馨
玄うげん	シ子ゲン	祝言
はうぐわん	ハ子グワン	判官
へうし	ヘ子シ	標帋、表紙
まうす	マ子ス	申
ようべ	ヨ子ベ	昨夜

二十一

Middle row: A list of words containing the character 子.

仮名遣
1891
http://dl.ndl.go.jp/info:ndljp/pid/862200/15
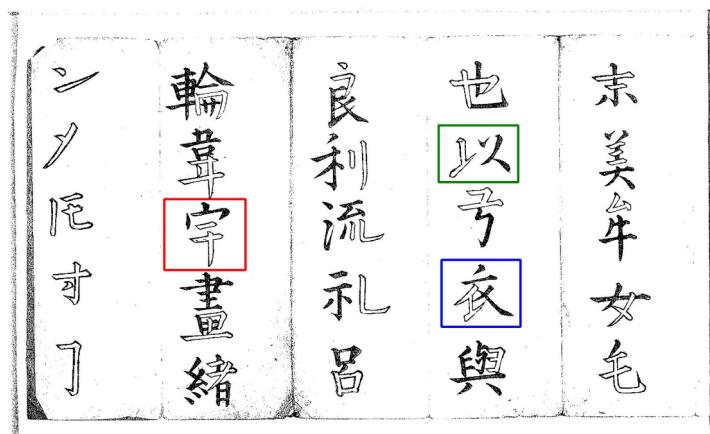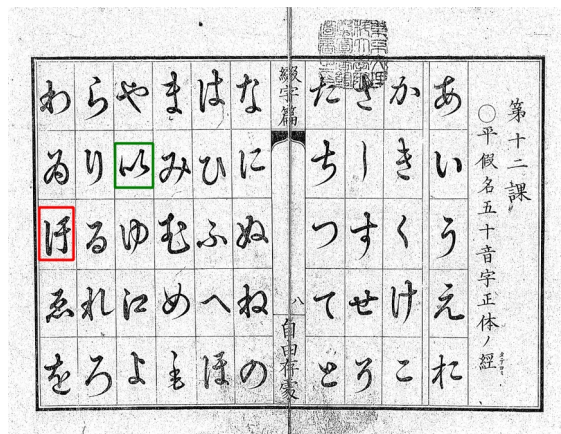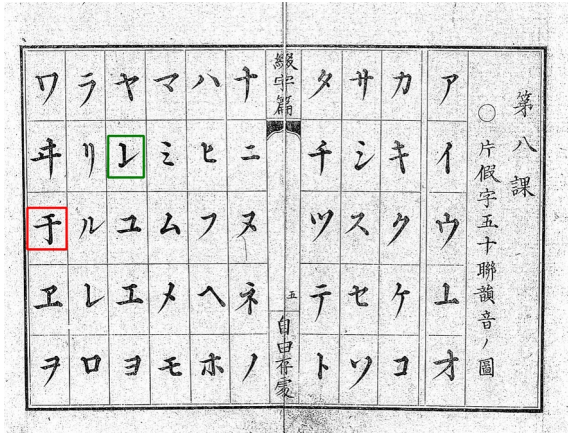
Red: the origin of the character 于
Green: the origin of the character ノ
Blue: the origin of the character ユ

Left image:	Right image:
片仮名元字	漢訳日語文法精義
1886	1906
http://dl.ndl.go.jp/info:ndljp/pid/862191	http://dl.ndl.go.jp/info:ndljp/pid/864026/13

Left chart:

和行	良行	也行	末行	波行	奈行	多行	左行	加行	阿行
ワ	ラ	ヤ	マ	ハ	ナ	タ	サ	カ	ア
ヰ	リ	[ノ]	ミ	ヒ	ニ	チ	シ	キ	イ
[于]	ル	ユ	ム	フ	ヌ	ツ	ス	ク	ウ
ヱ	レ	[衣]	メ	ヘ	ネ	テ	セ	ケ	エ
ヲ	ロ	ヨ	モ	ホ	ノ	ト	ソ	コ	オ

Right chart:

和行	畏行	也行	末行	波行	奈行	多行	左行	加行	阿行 (縦(音))	時	
ワ	ラ	(ヤ) マ	ハ	ナ	タ	サ	カ	(ア)	開	未來	過去 現在 過去 未來
(ヰ)	リ	(ノ) ミ	ヒ	ニ	チ	シ	キ	(イ)	承		
于	ル	(ユ) ム	フ	ヌ	ツ	ス	ク	(ウ)	轉		
ヱ)	レ	(衣) ノ	ヘ	ネ	テ	セ	ケ	(エ)	縦		
(ヲ)	ロ	(ヨ) モ	ホ	ノ	ト	ソ	コ	(オ)	合		

| 喉深音 | 舌巻音 | 喉淺音 | 唇重音 | 唇輕音 | 舌上音 | 舌頭音 | 舌牙音 | 喉牙音 | 喉正音 | 性質 |

Red: the character 于

Green: the character ノ

The YE in the chart above (in blue) is just a variant of the proposed 𛀀 (YE) with an extra stroke going diagonally to the bottom right. However, this version is rarely used. This variant shows that YE originated from 衣 and is not a ligature of イ and エ.

Left image:
仮名遣の栞
1892
http://dl.ndl.go.jp/info:ndljp/pid/862214/7

Right image:
国語学文典：普通教育
1894
http://dl.ndl.go.jp/info:ndljp/pid/863770/19

Left image:
Red: the character 于
Green: the character ㇄

Middle and right images:
Red: the character け
Green: the character (㇄) being used as YI as opposed to I.

綴字篇8課の1ページ目と2ページ目 1873	綴字篇第12課の1ページ目と2ページ目 1873	空欄だらけの50音図 **2019** https://fontring.web.fc2.com/works/articles/0002.html

Left image in red: the transitional form between 紆 and け.
Right image in red: the character 于.

仮名遣
1891
http://dl.ndl.go.jp/info:ndljp/pid/862200/

Middle row: A list of words containing the letter ン.
Top row: see that 𛀁 (𛀁) corresponds to ン (rather than い (I)).

仮名遣
1891

More proof of ん being used as "YI" rather than "I".

Left image:
仮名遣
1891
http://dl.ndl.go.jp/info:ndljp/pid/862200/8

Middle image:
日本大文典. 第1編
1897
http://dl.ndl.go.jp/info:ndljp/pid/864050/15

Right image:
仮名遣の栞
1892
http://dl.ndl.go.jp/info:ndljp/pid/862214/7

行乃以を

行の延とと定むべし。ゐ、宇の冠を取まるなり。宥は〔説文〕〔廣韻〕

行乃以を

阿行乃宥とを阿行乃衣とく耶。和行乃宇とウ。阿行乃宥とウ。和行の宇を

らず。いかでかかゝる理はあらむ。
又、かの僧空海の作なりと稱する、色波歌
十七音なりといふ説もあれど、そは深く
ものなり。我國の聲音は、もと、五十なりし
は和行也行の〔上〕｜｜〔于〕｜｜〔工〕と、その發音甚だ似
り、言語上、そを混同してもちゐたるもの
上通俗を旨として、製作せしものなれば、
一方の音を省きて、作りたるに過ぎず。さ

Red: the character 于
Green: the character レ
Blue: the character 工

Middle Top image:
明治役者絵版画帖
1900
http://dl.ndl.go.jp/info:ndljp/pid/993592

Middle Bottom image:
日本文典. 巻之上
1897
http://dl.ndl.go.jp/info:ndljp/pid/864063/5

Right image:
日本大文典. 第１編
1897
http://dl.ndl.go.jp/info:ndljp/pid/864050/11

Left image: ヶ used as a regular kana.
Both images: *You'on* (拗音) with 干.

Left image:
日本文典. 巻之上
1897
http://dl.ndl.go.jp/info:ndljp/pid/864063/8

Right image:
西字五十音図：一名・羅馬字五十音
1886
http://dl.ndl.go.jp/info:ndljp/pid/862390/5

You'on (拗音) with 于.
The YE in the chart above is just a variant of the proposed 亠 (YE) with an extra stroke going diagonally to the bottom right. However, this version is rarely used. This variant shows that YE originated from 衣 and is not a ligature of イ and エ.

国語学文典：普通教育
1894

Red: 于 being used to represent the syllable WU.
Green: ⺑ as "YI" as opposed to "I".

日本文典. 巻之上
1897
http://dl.ndl.go.jp/info:ndljp/pid/864063/6

Green: the characters レ and 𛀸
Blue: the character エ

単語独稽古
1874
http://dl.ndl.go.jp/info:ndljp/pid/863316/2

現行五十音圖となったのである。當の字形即ち[と]、[エ]を設けて、一完全なるものとなって、ヤ行イ、

Green: the character ヰ
Blue: the character エ

Left image:
日本語原
1916
http://dl.ndl.go.jp/info:ndljp/pid/1871505/104

Right image:
音図及手習詞歌考
1918
http://dl.ndl.go.jp/info:ndljp/pid/926012/12

Extras

Japanese wiki with a page on け/干:

https://www.wdic.org/w/CUL/%E3%82%8F%E8%A1%8C%E3%81%86%E6%AE%B5

Japanese Wikipedia pages:

https://ja.wikipedia.org/wiki/や行い
https://ja.wikipedia.org/wiki/や行え
https://ja.wikipedia.org/wiki/わ行う

Font for iPhones with け in it:

https://fontring.web.fc2.com/works/articles/0002.htm

Various sources plus the history of hiragana 紆→け:

https://kobunworld.blog.fc2.com/blog-entry-1.html

Glyph Wiki:

http://glyphwiki.org/wiki/hiragana-wu?view=all
http://glyphwiki.org/wiki/hiragana-small-wu?view=all
http://glyphwiki.org/wiki/katakana-wu?view=all
http://glyphwiki.org/wiki/katakana-small-wu?view=all
http://glyphwiki.org/wiki/katakana-vu
https://glyphwiki.org/wiki/katakana-yi
https://glyphwiki.org/wiki/katakana-small-yi?view=all
https://glyphwiki.org/wiki/katakana-ye
https://glyphwiki.org/wiki/katakana-small-ye

Last but not least, huge thanks to http://dl.ndl.go.jp/ for providing all the old scanned books.

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://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:** Proposal to Encode Missing Japanese Kana
2. Requester's name: Abraham Gross
3. Requester type (Member body/Liaison/Individual contribution): Individial Contribution
4. Submission date: 2020-01-05
5. Requester's reference (if applicable):
6. Choose one of the following:
   This is a complete proposal: ✓
   (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):
      Proposed name of script:
   b. The proposal is for addition of character(s) to an existing block: ✓
      Name of the existing block: Kana Extended-A (1B100-1B12F) and Small Kana Extension (1B130-1B16F)
2. Number of characters in proposal: 6
3. Proposed category (select one from below - see section 2.2 of P&P document):
   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 ____
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. Fonts related:
   a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? Abraham Gross
   b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): Abraham Gross abrahambgross@gmail.com
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)? Yes
   Sorting order information is enclosed

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.

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? — Japan National Body
      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? — Yes
   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: ....
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)? — No
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? — Yes
         If YES, is a rationale for its inclusion provided? — Yes
         If YES, reference: — In proposal
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? — Yes
         If YES, is a rationale for its inclusion provided? — Yes
         If YES, reference: — In proposal
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 characters? — No
    If YES, are the equivalent corresponding unified ideographic characters identified? ....
         If YES, reference: ....