The past, present and future of Hentaigana

Standardization for Information Interchange

Authors: TAKADA Tomokazu[[1]](#footnote-1), YADA Tsutomu[[2]](#footnote-2), SAITO Tatsuya[[3]](#footnote-3)

translation: Tatsuo Kobayashi (JNB),

Daniel Kobayashi-Better (Osaka University)

**Abstract**

Hentaigana are variants of hiragana; these variants are rarely used in the present day. However, Hentaigana were used for personal names until 1947, creating a need for a standardized coded character set of Hentaigana for administrative uses such as the basic resident register. There is also a demand for the ability to input and display Hentaigana on computers for academic purposes such as in the historical study of the Japanese writing system and studies of the history of Japan. Using characters collected from letterpress printing and digital fonts, we defined a set of Hentaigana to be used in the exchange of academic information. This set includes some variants that originate from the same Chinese character in order to accurately portray functional distinction in Hentaigana use. We plan to submit a proposal to add this set along with the Hentaigana used for administrative purposes to the ISO/IEC 10646 standard on October 2015.

1. Foreword

ISO/IEC 10646(Universal Coded Set) has already encoded almost all currently used modern characters and scripts, and has continued to encode historic scripts. Hentaigana is one of such new challenge. In 2009, a Dutch expert submitted a proposal to add Hentaigana in UCS, from the view point of cultural heritage and Orientalism.

In Japan, there has been demand to some extent for using Hentaigana in the digital environment since mainframe era, and vendors have been holding some Hentaiganas as private use characters. The main use case for Hentaigana is the governmental treatment of personal names (mainly first name of females), and the requirement to standardize Hentaigana in UCS is arising from this use case.

Apart from the governmental usage and cultural heritage, there are requirements to standardize Hentaigana encoding in academic work, especially historical studies of Japanese characters and its writing system, and transcriptions[[4]](#endnote-1) of Japanese historic documents for precise research. However, there is no stabilized set of Hentaigana yet. This survey is to establish a stabilized set of Hentaigana for these purposes.

In this paper we discuss the following

1. What is Hentaigana
2. A brief history of Hentaigana and encoding challenges
3. The process of selecting a Hentaigana set

２．What is Hentaigana

In the Japanese modern writing system, concerning Hiragana and Katakana usage, only one Hiragana letter corresponds to each phonetic value, with a few exceptions of ha/wa, e/he, o/wo, or long vowels.

However, this writing method came into widespread use very recently, considering the long history of Hiragana usage from 9th century. Around a thousand years until the end of 19th century, it was common for several letters to correspond to each phonetic value. The word “Hentaigana” is used two ways in the historic research of Japanese characters and its writing system. One is the writing system itself with multiple letters for one phonetic value. The other refers to any Hiragana letters that are not in common use. In this discussion, we will use the word “Hentaigana” in the latter meaning.

After 1900 A.D. (Meiji 33), Hentaigana faded out from the daily use in the Japanese writing system.

In this year, 1900 A.D. the Ministry of Education promulgated 「教授用仮名及び字体」(Kana for Use in Education) in 「小学校令施行規則」（文部省令第14号）(Ordinance for Enforcement of the School Education Act (Ordinance of the Ministry of Education No.14)), which defined the Hiragana letters to be used in education. The hiragana letters defined in this promulgation are still in common use today.

As a matter of fact, this promulgation was only focused on elementary education, and the promulgation revoked in 1908 along with the writing rule of long vowels in on-yomi words using ー, so there are no explicit law or ordinances that explicitly prescribe the use of Hentaigana in common Japanese, except for in persons’ names.

However, even before promulgation regarding education, after Meiji 20, with the spread of movable-type printing, the use of Hentaigana had declined rapidly and the promulgation only confirmed the actual usage. In this way, the abolishing of this promulgation made no effect either in education or society in general, and the usage of Hentaigana faded away from the Japanese writing system. Currently, there is no disadvantage for anyone who lacks Hentaigana literacy. However, the use of Hentaigana has not completely disappeared from Japanese documents.

There still has been very few use case of Hentaiganas, for example Yukio Mishima, a famous postwar generation novelist, used a few Hentaigana in his manuscripts. The use of Hentaigana for persons’ names, especially female names, is another vestige of the Hentaigana generation.

In addition, in the modern landscape, we can sometimes find uses of Hentaigana, in store names and logos. In this sense, Hentaigana is still alive in Japanese cultural scene.

That Japanese writing system finally changed to the modern Hiragana system (i.e. one tone value is represented by one Hiragana letter), is very reasonable in regard to the economics of language. If so, why was Hentaigana present in the writing system so long—more than one millennium? We will discuss this issue below.

The origin of Hentaigana is Man’yogana. Man’yogana is derived from Kanji ignoring their meaning, only inheriting their sound and shape, and are used to transcribe the Japanese language. When used as Man’yogana, the meaning of Kanji is abandoned, leading to the disappearance of any difference between Kanji with the same or similar sound. There had been no strict rule to select one letter among many possible choices before the modern era. Man’yogana had many Kanji letters for each sound value. Hiragana inherited the characteristics of Man’yogana and had multiple letters for each tone value. As a matter of fact, there was a tendency in Man’yogana itself of a reduction in the variety of the Kanji letters for each sound value during the expansion of its area of use. However, before this reduction in variety, Hiragana itself developed shape modifications to make them visually different to Kanji.

Also, the fact that society in the Heian era, when Hiragana was developed, was relatively stable also contributed to the situation of Hiragana. Generally speaking, changes in the structure of writing system occur along with changes in society. With the increase of the population with literacy as social demand, a more rational writing system was developed for the new literate people. However, society in the Heian era remained stable for a long time, and there was no explicit development to reduce the letters for each phonetic value, rather they developed aesthetically. Hiragana letters were not reduced but increased in the Heian era.

In the case that there is more than one letter for each phonetic value, it is rational for people to create functional difference between them. Also, it is very natural to imagine people when writing would require some guidelines on how to use multiple letters for each phonetic value appropriately. There are many studies in historical research on the Japanese writing system about the functional differences among multiple letters for one phonetic value.

Examples are as follows:

(1) specific letters used in specific words

「今日(けふ)」→「 C:\Users\takada\Desktop\学術用_png_150317\gh0540.pngふ」

「参らせ」→「C:\Users\takada\Desktop\学術用_png_150317\gh2050.png いら C:\Users\takada\Desktop\学術用_png_150317\gh0890.png」

(2) different letters used for their position in a specific word (top, middle or bottom), or for particles.

〈カ〉か－語頭／ C:\Users\takada\Desktop\学術用_png_150317\gh0270.png－語中語尾（2）

〈シ〉 C:\Users\takada\Desktop\学術用_png_150317\gh0770.png－語頭／し－語中語尾（3）

1. C:\Users\takada\Desktop\学術用_png_150317\gh1750.png－語頭／は－語中語尾／ C:\Users\takada\Desktop\学術用_png_150317\gh1670.png－助詞（4）

(3) different letters for the same phonetic value appearing in two adjacent lines

In addition, after the Middle Ages, we find examples where different letter shapes are used to distinguish differences in phonetic value.

(4) The difference of phonetic value is expressed with different of letter shapes.

〈チ(清音)〉ち／C:\Users\takada\Desktop\学術用_png_150317\gh1090.png〈ヂ(濁音)〉 （5）

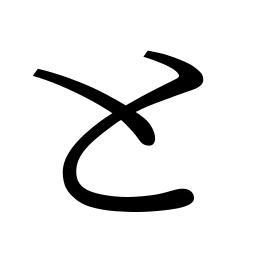
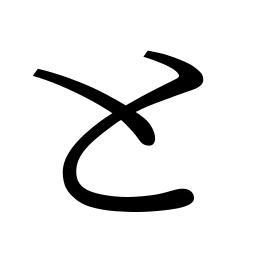
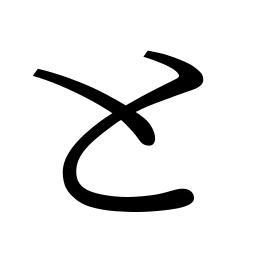
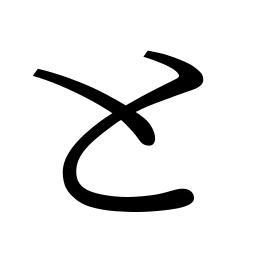
〈ユ(直音)〉ゆ／C:\Users\takada\Desktop\学術用_png_150317\gh2410.png〈ユ(拗音)〉 （6）

〈ツ・ヅ〉つ・C:\Users\takada\Desktop\学術用_png_150317\gh1160.png ／〈促音〉（7）

After the use of Hentaigana had disappeared from common use, reprints of classical literature with movable type usually used only modern Hiragana. However, considering the functional usage of different letter shapes, there is possibility of losing very important information from original manuscript if only modern Hiragana is used in transcription. Moreover, looking at the example in (4), we can find that the functional difference of different letter shapes occurs not only between characters originating from two different Kanji, but also between two different letter shapes originating from the same Kanji.

The next example is a humorous anecdote from the Edo era. In modern sense, we can not find the difference, however, to completely understand the anecdote itself, we need to distinguish the difference of two letter shapes that represent the sound /to/. Punctuation marks are added by the author of this paper.

Note by translator: Very sorry without English translation, but it is very similar difficulty to translate James Joyce’s “Finnegans Wake” into Japanese.

　・大和国よしのあたりに、むすめ弐人持ける人、てうあひのあまり、姉の名をしきぶ付、妹をちぎりよびける。ある人のいはく、「こなたの姉娘しきぶはきこへましたが、いもふをちぎりはいかゞ」たづねければ、「をのをのは謡御ぞんじない見へました。それはもんもう也。井筒のうたひにありありしるしおかれた、紀のありつねの娘はならびなきびじんゆへ、なり平心をつくされしよし、我等がむすめもあつぱれおるまじきぞんじよりにてちぎり付申侯」いふ。「して其井づゝのうたひは」へば、しさいらしく、「その比は紀のあり常の娘ちぎりうたひます」いはれた。（『軽口御前男』）

In the lyrics of 謡曲「井筒(izutsu)」, the text is 「紀有常の娘と契り」, however, the scatterbrain misunderstood the text to 「紀有常の娘『とちぎり』」i.e. 『とちぎり』 as the lady’s proper name. This anecdote has four 「とちぎり」strings. The first three times use the Hentaigana letter 「止」to suggest that「とちぎり」is a proper name, and the last time, uses 「と」to declare that 「とちぎり」means to have a love affair.

This is a very adequate example that in the pre modern Japanese writing system, sometimes letter shape differences carried very important functionality to enable rich exchange of information.

３．Hentaigana and coded character sets

Japanese coded character sets, such as JIS X 2028, JIS X 0212 and JIS X 0213, have no code positions for Hentaigana. However, in the 1970s, when he first Japanese first coded character set JIS X 0208 (was named JIS C 6226 when published) was published, Japanese main frame vendors treated Hentaigana as part of their private character set. In vendors’ codebooks, Fujitsu includes 168 letters (FUJITSU JEF, 1980), NEC includes 152 letters (『日本電気標準文字セット辞書＜拡張＞』, 1983), and Hitachi includes 152 letters (『HITAC文字パターン辞書／コードブック』,1987).

These Hentaigana are prepared to handle human proper names in practical works.

The handling rule of Hentaigana in personal names is included in戸籍法施行規則 (Ordinance for Enforcement of the Family Register) Article 60 enacted on 1st January 1948. Article 60 is the rule to limit characters to be used for newly born children’s names. Concerning Kana, after 1st Jury 1948, Katanana or Hiragana (excluding Hentaigana) were permitted for use. In accordance with this rule, there had been no rule to prohibit Hentaigana for personal names before 31st December 1947. A person who has Hentaigana in his/her name and was born in 1947 would be 67 or 68 years old in 2016. Article 5 of戸籍法施行規則 (Ordinance for Enforcement of the Family Register) also regulates to keep records of deceased or missing persons for 150 years. Accordingly, governance related to the family register needs to keep using Hentaigana around 200 years.

The Ministry of Justice regulated the Character/glyph set for the digitalized family registration system called戸籍統一文字 (Unified Character Set for the Family Registry) in法務省民一第928号民事局長通達 (Notice No. 928, issued by 1st Section, Civil Affairs Bureau, Ministry of Justice) promulgated on 1st April 2004.

In戸籍統一文字, 168 Hentaigana letters are included. This repertory is very similar to the repertories included in vendor systems, and it can be inferred that this 戸籍統一文字 repertory inherits the Hentaigana used in mainframes. This fact raised the requirement to standardize Hentaigana in UCS.

In JIS, there once had been discussion about the inclusion of Hentaigana. In commentary for JIS X 0213:2000 (pp.512) there is some mention about Hentaigana.

*4.4.5 Hentaigana Hentaigana is used in some areas, such as in text books for Japanese calligraphy, and there existed a small demand for standardization. However, it is very difficult to fix a set repertory of Hentaigana. Also it is very difficult to define stable glyph shapes from actual use cases that can be identified as graphic characters, and it is very difficult to clearly distinguish between Hentaigana and the handwritten shape of their original Kanji. For the above reasons, the committee finally decided not to include Hentaigana.*

From this description, it can be understood that in order to standardize Hentaigana as coded characters, it is necessary to settle representative glyph shapes for each Hentaigana letter.

On the other hand, there was a proposal to include Hentaigana in UCS from an individual Dutch expert in February 2009. This proposal is based on a Kanji searching software “Konjaku Mojikyo” and it includes 213 Hentaigana letters. In that occasion, JNB for ISO/IEC JTC1SC2 was requested to make comment for this proposal. The Japanese answer was as follows:

In Japan, there is still a requirement to standardize Hentaigana in coded character set standards especially in governmental use for personal names, however, the policy for sophisticated e-governmentization is still under discussion. This Japanese comment suspended the Hentaigana discussion in SC2.

４．Research and selection of Hentaigana set for academic purpose

Generally speaking, a character set for standardized coded character set should be sufficiently stable and fairly acceptable for a large number of people to use. This is the same for the historical study of the Japanese writing system and history of Japan (古文書学). Especially, it is very important to reach an agreement on a stable glyph shape as a graphic character between researchers, and for the repertory to be sufficiently usable for the purpose of historical study of the Japanese writing system and history of Japan.

To attain this goal, the selection of Hentaigana began with gathering Hentaigana letters used in modern letter press printing and digital fonts, and consulted dictionaries focused on Hentaigana[[5]](#endnote-2).

Usually, we would imagine old handwritten shapes by famous masters when thinking of Hentaigana shape. However, handwritten shapes have considerable variations and it is very difficult to consolidate representative shapes. Accordingly, we started with letter press printing and digital fonts, which have more stable shapes to aid selection.

The gathered sets are as follows:

(1) 精興社変体仮名　134字

(2) 中西印刷変体仮名　154字

(3) 築地活版五号明朝体活字変体仮名（明治27年）〔府川充男『聚珍録　第三編仮名』（三省堂、2005年）より〕　115字

(4) 教科書活字（明治25年・明治27年）〔板倉雅宣『教科書体変遷史』（朗文堂、2003年）より〕　明治25年：49字　明治27年：50字

(5) 今昔文字鏡変体仮名　380字

(6) Koin変体仮名〔http://www10.plala.or.jp/koin/koinHentaigana.html〕　335字

(1) and (2) are the Hentaigana sets which the printers of 『大日本古文書』 and『大日本古記録』[[6]](#endnote-3)use. These sets seem to cover the requirements for the study of the history of Japan. (3) was published in 1894, the period when the design of letter press typefaces became stabilized. (4) was used in elementary school text books in the Meiji era, and seems to display a fundamental set of Hentaigana. (5) and (6) are popularly used by researchers, in the present day. These sets contain Hentaigana letters with the voiced and semi-voiced sound marks. However, in our Hentaigana set, these voiced and semi-voiced letters have been omitted because the needs of these letters for scholarly purpose are very low.

Note by tlk: In Japanese proposal to SC2, theoretically these voiced and semi-voiced Hentaigana letters can be rendered with UCS architecture and functionality in general.

More over, we added Hentaigana letters appearing in12 previous studies on Kana character[[7]](#endnote-4) and Hentaigana letters used by 東京大学史料編纂所 (Historiographical Institute, the University of Tokyo).

After gathering the candidate Hentaigana letters, grouping these letters according to their similarity in shape, we then, checked these with Hentaigana dictionaries. Hentaigana dictionaries referred to are as follows:

(1) 『変体仮名の手引（改訂増補25版）』（中野幸一編、武蔵野書院、2009年）

(2) 『字典かな（新装版）』（笠間影印叢刊刊行会編、笠間書院、2003年）

(3) 『かな名跡大字典』（筒井茂徳編、角川書店、1981年）

(4) 『かな連綿字典』（竹田悦堂監修・佐野榮輝編、雄山閣出版、1990-1992年）

(5) 『用例かな大字典』（中田易直他編、柏書房、1978年）

All dictionaries have source information and text, therefore with this checking process, we could confirm the evidence of each Hentaigana.

With this process, we consolidated a 264 Hentaigana glyph repertory for academic use as follows:



Fig.1 The Hentaigana set for scholarly information exchange

The Hentaigana set for scholarly use consists of 47 phonetic values, 215 mother Kanji characters, 264 glyph shapes, and is arranged with three layers i.e. phonetic value, mother Kanji character and glyph.

The first layer, phonetic value does not reflect the historical change of phonetic values, but simply reflects the current phonetic values corresponding to modern Hiragana letters.

The second layer of mother Kanji characters, shows original Kanji from which Hentaigana is generated by modification to a cursive style. Usually, Hentaigana letters are displayed with their original Kanji characters, so the table contains the original Kanji character with corresponding Hentaigana. Accordingly, one phonetic value corresponds to multiple original Kanji characters. The ordering of the original Kanji characters is by their radical and stroke count.

The third layer is “glyph”. Hentaigana sometimes has different glyphs origination from the same original Kanji character. Referred in section two, these different glyphs with one original Kanji character have different functions in the writing system. Hence, in this Hentaigana set, needed different glyphs with the same original Kanji character have been included as necessary (ex. For mother Kanji「可」,「C:\Users\takada\Desktop\学術用_png_150317\gh0270.png」and「C:\Users\takada\Desktop\学術用_png_150317\gh0280.png」are correspond).

Also, classical texts sometimes contain one original Kanji character or one glyph used for multiple phonetic values. For these cases, in this table, we displayed only one phonetic value. For example, Hentaigana「C:\Users\takada\Desktop\学術用_png_150317\gh2370.png」is used for /YA/ and /YO/, but in this table only /YA/ is displayed. This issue should be discussed and resolved through the process of standardization. Cases that one glyph used for multiple phonetic values are as follows:

惡（C:\Users\takada\Desktop\学術用_png_150317\gh0030.png）：あ、を

家（C:\Users\takada\Desktop\学術用_png_150317\gh0310.png）：か、け

期（C:\Users\takada\Desktop\学術用_png_150317\gh0620.png）：こ、き

徒（C:\Users\takada\Desktop\学術用_png_150317\gh1160.png）：つ、と

等（C:\Users\takada\Desktop\学術用_png_150317\gh1360.png）：と、ら

而（C:\Users\takada\Desktop\学術用_png_150317\gh1520.png）：に、て

子（C:\Users\takada\Desktop\学術用_png_150317\gh1560.png）：ね、こ

馬（C:\Users\takada\Desktop\学術用_png_150317\gh2250.png）：め、ま

夜（C:\Users\takada\Desktop\学術用_png_150317\gh2370.png）：や、よ

无（C:\Users\takada\Desktop\学術用_png_150317\gh2190.png）：む、も、ん

Note by translator: JNB proposal resolved this issue, with assigning different identifier for the pairs of phonetic value and original Kanji character.

５．Future plan

Japan National Body for ISO/IEC JTC1/SC2 is going to submit the consolidated proposal with the 168 governmental use set and an additional 118 academic use set, making a total of 286 Hentaigana letters, to be standardized in ISO/IEC 10646.

References

1) 矢田勉. 国語文字・表記史の研究. 汲古書院. 2012. 825p.

2) 迫野虔徳. 定家の「仮名もじ遣」. 語文研究. 1974, no. 37, p. 39-46.

3) 安田章. 仮名文字遣と国語史研究. 三省堂. 2009. 312p.

4) 伊坂淳一. 藤原俊成の用字法・試論―自筆本『廣田社歌合』における機能的用字法―（Ⅰ）（Ⅱ）. 学苑.

1988, no. 577, 578, p. 179-189, 59-71.

5) 鈴木真喜男「. 地」のかな―定家自筆本における―. 国語研究. 1958, no. 8, p. 59-65.

6) 長谷川千秋. 直音と拗音を書き分ける仮名文字遣. 国語国文. 1997, vol. 66, no. 7, p. 36-52.

7) 蔵野嗣久. 天正狂言本の「つ」の仮名について. 安田女子大学紀要. 1973, no. 3, p. 39-43.

8) 日本工業標準調査会. 7ビット及び8ビットの2バイト情報交換用符号化拡張漢字集合 JIS X 0213:2000. 日

本規格協会. 2000. 541p.

9) 平本健二. 電子行政における文字環境の整備. 情報管理. 2015, vol. 57, no. 11, p. 799-808. http://doi.

org/10.1241/johokanri.57.799, (accessed 2015-07-08).

10) 小林龍生. 字体と字形のはざまで：文字情報基盤整備事業を例として. 情報管理. 2015, vol. 58, no. 3, p.

176-184. http://doi.org/10.1241/johokanri.58.176, (accessed 2015-07-08).

1. National Institute for Japanese Language and Linguistics [↑](#footnote-ref-1)
2. Osaka University [↑](#footnote-ref-2)
3. Senshu University [↑](#footnote-ref-3)
4. In this paper “transcription” means to rewrite texts in manuscripts and wood cut printings into modern style texts using modern character shape and design. Sometimes, through adding punctuation marks and replacing Hiragana strings into Kanji, transcripter’s interpretation would be included. [↑](#endnote-ref-1)
5. This project was conducted as the research project 「文字環境のモデル化と社会言語科学への応用」(from Oct. 2009). The members are銭谷真人（国語研）,小助川貞次（富山大）,當山日出夫（立命館大）,Yada, Saito, Takada. [↑](#endnote-ref-2)
6. Hentaigana is used in transcription of 和歌(waka) and 仮名消息(Kanashousoku). [↑](#endnote-ref-3)
7. 池田亀鑑『古典の批判的処置に関する研究』（岩波書店,1941年）,山田俊雄『日本語の歴史 別巻 言語史研究入門』（平凡社,1976年）,浜田啓介「板行の仮名字体―その収歛的傾向について―」（『国語学』118,1979年）,築島裕『日本語の世界 5 仮名』（中央公論社,1981年）,玉村禎郎「『春色梅兒誉美』における仮名の用字法」（『国語文字史の研究 2』,和泉書院,1994年）,松下なるみ「平仮名の字体と字源」（『漢字百科大事典』,明治書院,1996年）,久保田篤「『浮世風呂』の平仮名の用字法」（『成蹊国文』30,1997年）,内田宗一「黄表紙・洒落本の仮名字体」（『国語文字史の研究 4』,和泉書院,1998年）,内田宗一「『偐紫田舎源氏』の仮名字体―作者自筆稿本と板本の比較考察」（『待兼山論叢文学篇』32,1998年）,内田宗一「馬琴作合巻『金比羅船利生纜』の仮名字体」（『国語文字史の研究 5』,和泉書院,2000年）,久保田篤「江戸板本の表記の多様性―洒落本『傾城買二筋道』の場合」（『成蹊国文』42,2009年）

   [translator’s note]

   This article is originally carried in “Journal of Information Processing and Management” 2051 vol.58 no.6 published by Japanese Science and Technology Agency (JST). [↑](#endnote-ref-4)