

**Date:** July 22 2015  
**Title:** Request for Comments on HENTAIGANA proposal  
**Source:** Japan  
**Action required** For review  
**No. of pages:** 4(this document),  
23(attachment of HENTAIGANA repertory)

## Request for Comments on HENTAIGANA proposal

JNB developed a proposal of standardization of HENTAIGANA.

The source of the repertory is Moji-Joho-Kiban project led by Japanese government. This repertory consists of 299 HENTAIGANAs.

JNB proposes to assign Standardized Variation Sequences in UCS to encode these Japanese HENTAIGANAs. We request members and experts of SC2 to comment on its encoding scheme.

As for the repertory of HENTAIGANA, JNB encourage them to enter comment to the following public review site. Sorry that it is Japanese site. Please contact us if you need any help on this site.

<http://mojikiban.ipa.go.jp/4436.html>

Attached is the detail proposal.

## Proposal for Japanese HENTAIGANAs

Japanese phonograms, widely known as “HIRAGANA” today, had been derived from handwritten shapes of ancient Hanzi Ideographs. The meanings of the original Ideographs (“mother ideographs”) are completely eliminated and they only carry sound.

In year 1900, Japanese government selected one phonogram for each phonetic value and announced not to use other phonograms in elementary education. Afterward, the selected phonograms are called “HIRAGANA” and others are called “HENTAIGANA”, the meaning is variants of a HIRAGANA. HIRAGANAs already have code points in UCS, in the HIRAGANA Block.

Today, the use cases of HENTAIGANA are very limited, but they are still actually used especially in persons’ names.

They are often used in the first names of female persons born before 1948<sup>1</sup>. It is required by law in the Japanese family registration system to preserve records of persons for 150 years after they have passed away, so HENTAIGANAs used for these persons will still be needed for more than 150 years in administrative areas.

The other use cases are in scholarly studies, especially research on the history of Japanese literature, writing system, and transcriptions of historic documentations.

In both areas, processing HENTAIGANA in a digital environment is desired and, in many case, HENTAIGANAs are used assigned within private use area / plane in the UCS code space.

For enhancing interoperability of IT systems including e-government systems and fostering a fair competitive and open procurement environment, international standardization of HENTAIGANA cannot be avoided.

The attached chart is the repertory of HENTAIGANA which reflect

---

<sup>1</sup> The Japanese government explicitly excluded HENTAIGANA from the character list, which can be used for newly born children’s first names, in 1948.

administrative and academic needs.

#### [Sources of the repertory]

Mainly there are two different areas of sources.

One is the officially announced and used list for the Japanese family registration system. There are 168 HENTAIGANAs from this source.

The other is scholarly study. The Japan NB and several voluntary scholars gathered and consolidated HENTAIGANAs mainly used for printing purposes after the Meiji era<sup>2</sup>, and checked with several HENTAIGANA dictionaries containing examples of actual use cases. There are 275 HENTAIGANAs from this source.

The Japan NB consolidated these two different sources and made a repertory containing 299 HENTAIGANAs.

#### [Encoding model]

The Japan NB proposes to assign standardized variation sequences (SVS) to HIRAGANAs to specify HENTAIGANAs as a variation of corresponding HIRAGANAs.

Basically, a HENTAIGANA is identified by the pair of its phonetic value and its mother ideograph. That is, basically, one SVS will be assigned to a unique phonetic value - mother ideograph pair. But if there is an actual use case that more than two HENTAIGANAs which have the same phonetic value - mother ideograph pair are intentionally used differently, then different SVSs are assigned.

Note1. It is preferable to assign the first variation selector to specify the HIRAGANA itself.

Note2. There possibly are numerous (infinite) shape variants for one HENTAIGANA because the shape of HENTAIGANA came from handwritten

---

<sup>2</sup> 1868A.D.~1912A.D.

style of ancient Hanzi Ideographs and the level of simplification can vary. So the Japan NB proposes a single variation sequence for each pair of phonetic value and mother ideograph. Visual representations of HENTAIGANAs in this proposal are only examples for rendering, and they can be rendered in a wide variety of ways.

Note3. There are a few exceptions in this Japanese proposal to assign more than one variation sequence for one pair of phonetic value and mother ideograph. There are two reasons for this. Firstly: There are HENTAIGANA shape variants that the Japanese government have already officially assigned and used in the actual working systems. Secondly: In the history of the Japanese writing system, there are some functional differences corresponding to the degree of simplification of the mother ideographs.

Note4. There are a few cases that one mother ideograph is associated with plural phonetic value. In this case, a different SVS will be associated with a very similar or identical visual representation.

Note5. Some HENTAIGANAs may be used with a voiced sound mark or semi-voiced sound mark. For such cases, COMBINING KATAKANA-HIRAGANA VOICED SOUND MARK (U+3099) and COMBINING KATAKANA-HIRAGANA SEMI-VOICED SOUND MARK (U+30A0) may be used with HENTAIGANAs.

According to section 16.6.1, 4.14 and 20.5 of ISO/IEC 10646:2014, the order of characters should be “base character (ie. HIRAGANA) + VS + combining character (ie. U+3099 or U+30A0)”, but a more clear explanation or reminder of character order will be needed.