

**Title:** Status of hentaigana proposal

**Source:** Ken Lunde (Adobe)

**Status:** Individual contribution

**Action:** For consideration by UTC

**References:** [L2/15-343](#) (aka WG2 N4708), [L2/16-037](#), [L2/16-040](#) & [L2/16-053](#)

**Date:** 2016-04-28

This document provides the current status of Japan's *hentaigana* (変体仮名) proposal, in the form of an exchange between Shuichi TASHIRO (田代秀一) and myself, along with a response from Dr. Nicolas Tranter.

The first email was sent by me to Tashiro-san on 2016-02-06, which was intended to encapsulate all of the UTC's comments to date:

Tashiro-san,

Please accept my apologies for the delay in relaying to you all of the feedback on Japan's latest hentaigana proposal (aka L2/15-343):

<http://www.unicode.org/L2/L2015/15343-hentaigana-revised.pdf>

As you are aware, UTC #146 took place a week ago, and one of the discussion topics was this latest proposal. Please find attached the following four documents:

15343-hentaigana-rev-spreadsheet-annotated.xlsx  
16037-script-rec.pdf  
16040-hentaigana-unification-candidates.pdf  
hentaigana-lunde-tranter.pdf

Please share these documents and this email with others in Japan who are involved with this important proposal.

The first document, 15343-hentaigana-rev-spreadsheet-annotated.xlsx, is simply an annotated version of the spreadsheet that was supplied with L2/15-343, which differs only in that it highlights hentaigana characters that share the same parent ideograph (kanji) for easier identification. (Note that two different highlight colors, yellow and orange, are used only because some such cases are contiguous, and a different highlight color acts as a visual separator.) It would be useful if this spreadsheet were to be prepared in this fashion, because it is helpful to easily see which hentaigana characters share the same parent ideograph.

The second document, 16037-script-rec.pdf, is the Unicode Script Ad Hoc recommendations to the UTC, and their recommendation for Japan's latest hentaigana proposal can be found on pp 8 and 9, and has been highlighted for convenience. The UTC did briefly discuss HENTAIGANA LETTER E-1 versus U+1B001 HIRAGANA LETTER ARCHAIC YE, and the main concern is that U+they differ only by name, and that unification seems like the prudent thing to do. An alias, such as "hentaigana letter e-1," can be added to U+1B001 to make the unification more easily understood by implementers.

The third document, 16040-hentaigana-unification-candidates.pdf, is my own attempt at determining which hentaigana characters could potentially be unified, based on my (perhaps naïve) understanding of calligraphic rules and practices. As noted in the document, these candidates are limited to hentaigana characters that have only an

Academic Use source reference, and for which there exists one or more corresponding hentaigana characters that share the same parent ideograph and that have a Family Registration source reference.

The fourth document, hentaigana-lunde-tranter.pdf, is an email exchange between myself and Nicolas Tranter, who is a hentaigana expert. You and others were included in that email exchange, but it is useful to keep this record. I don't fully understand all of the points that Nicolas raised, and therefore defer to the hentaigana experts in Japan to consider them. The last page of this document includes a short list of possible hentaigana/hiragana unification candidates. Again, the hentaigana experts in Japan should take this into consideration for the next version of the proposal.

In closing, the UTC would like to thank those in Japan who have been working on this important proposal for their continued patience, and for giving the UTC and others an opportunity to provide feedback. Please note that there is no overall objection to Japan's latest proposal, but it is nonetheless important for Japan to consider all of the recent feedback when preparing the next version of this proposal.

With best regards...

-- Ken Lunde

## I received the following informative reply from Tashiro-san on 2016-04-14:

Dear Ken Lunde-san,

Sorry for late response.

I appreciate UTC's efforts for reviewing proposal of Japan on Hentaigana standardization including getting contribution from global experts of Japanese language.

I recognized that UTC has very positive position to Japanese proposal with minor modification (L2/15-343 or SC2/WG2 N4708) in accordance with US/UTC comments in SC2 Matsue meeting. However, there still remain a couple of questions and comments to be considered.

I forwarded your e-mail dated February 06 to Japanese Hentaigana experts and discussed about each issue.

The answers I got from the experts are as following.

The first issue is about a group of Hentaiganas which Dr. Nicolas Tranter mentioned as "kanji used phonographically". He mentioned that these may not be Hentaigana but cursive images of Kanji characters, with remarking he is an expert of manuscripts in late Edo era but not an expert of that in medieval era.

Japanese Hentaigana experts kindly showed evidence for Hentaiganas in this group.

Attached file "Evidences\_against\_kanji\_used\_phonographically.xlsx" includes the evidences.

The images of the evidences are from Japanese short poems, called WAKA. In Japanese literature tradition, WAKA had been written in Kana letters only, and no Kanji characters. So, such evidences are good enough to decide that these characters are not Kanji character but Kana letters.

# Please understand that it is strictly prohibited to copy or transfer the content to other media (printed materials, broadcasting, websites, etc.) without the explicit permission of the National Institutes for Cultural Heritage. So we only can exchange the URLs of them.

The second issue is about pairs and triples of Hentaiganas, which share identical pairs of mother ideograph and phonetic value.

In general, Japan's proposed model of coding is to assign character code to a pair of mother ideograph and phonetic value, and the associated image is informative, in principle. However, there remains some exceptions. That is, we assigned plural character codes to some identical pairs of mother ideograph and phonetic value.

In the case that academic use Hentaiganas share identical pairs of mother ideograph and phonetic value, there are obvious reason such as the Hentaiganas in the pairs or triples have functional difference as Prof. Yabe presented in Matsue meeting. "EMIC" Dr. Nicolas Tranter mentioned.

In the case of Hentaiganas which appear only in family registration use share an identical pair of mother ideograph and phonetic value, there is no room to argue because these Hentaiganas has been actually used as differentiated characters in governmental IT systems to identify persons' names.

In the case that a Hentaiganas is used only in family registration use and another Hentaiganas is used only in academic use and these Hentaiganas share an identical pair of mother ideograph and phonetic value and identified as different characters, Japanese experts decided not to unify them.

Japanese experts tried to find actual use cases of such Hentaiganas as used only in family registration use from manuscripts of academic field with best effort, but could not find any actual use cases.

In another word, Japanese linguistic experts could not find the evidences why these glyphs are included in family registration use characters, with academic research procedure.

Accordingly, Japanese experts and JNB have no clear rationale to unify or not to unify these pair of glyph images.

With these consideration, Japanese experts and JNB decided that in these cases, it is safer not to unify these pairs and triples. There are implicit possibilities that some new evidences for different functionalities will be found in future.

The example of the pairs or triples in this case is shown in "FoundNoEvidence.xlsx"

As for "E-4" in this file, Japanese experts had reviewed this Hentaigana, and once deleted it from their academic proposal. I added the academic ID of this Hentaigana in the file.

As for 1B001 Archaic Ye, we are discussing to propose to add new Character Name Alias "HENTAIGANA LETTER E-1" to U+1B001 as normative information and informative information about mother ideograph. And to propose to change the corresponding glyph image to the glyph of E-1 in our current Hentaigana proposal (JTC 1/SC 2/WG 2 N4708).

And we will withdraw E-1 from the proposal.

Best regards,

Shuichi Tashiro, IPA

-----  
Shuichi TASHIRO | General Manager, Open Standards Promotion Center  
tashiro@ipa.go.jp | Information-technology Promotion Agency (IPA)  
tashiro@acm.org | http://www.ipa.go.jp/ tel: +81-3-5978-7507  
-----

Please see the two Excel files that are attached to this document, which are referenced in Tashiro-san's 2016-04-14 email.

Dr. Nicolas Tranter, who has reviewed previous versions of Japan's hentaigana proposal, and who is an expert in this area, had an opportunity to review Tashiro-san's 2016-04-14 response as encapsulated in this document, to include the two Excel files that are attached. The following is his feedback, received on 2016-04-26:

Dear Ken,

Thanks for the document L2/16-085. I've read through it and looked at the two attachments, particularly considering Professor Tashiro's.

I was always broadly in favour of the Japanese hentaigana proposal, and my main concern was the odd choice of hentaigana by the Ministry of Justice. This, as we all agree, is already official within the ministry and therefore is not negotiable. It's pleasing to see Prof Tashiro's team agreeing with me on the lack of academic support for some of the ministry's decisions.

I'm content with the rest of the chart. I think Prof Tashiro's evidence of kanji-free waka poetry is very persuasive, and since the example he gives of TE-4 is the oldest MS of this major (indeed, paramount) member of the Classical Japanese poetry canon I accept his argument.

My research interest ultimately is gaining the facility of a Unicode-encoded font, and I personally don't have any further objections. I wish everyone all the best on what I hope will be the final stages of the proposal. I wonder, which future version of Unicode is hentaigana likely to be included in, everything else going well?

Thanks for calling me a hentaigana expert! My specialisms are writing systems generally and Classical and Late Edo Japanese, both of which give me a strong vested interest in this. I've published one book chapter using hentaigana with jpegs (with a couple of publisher errors mixing up NA-1 and NI-2 as a result), and I've another chapter using hentaigana coming out next year. But I can't really be called an expert. More a well-informed amateur.

All the best, Nic.

Dr Nicolas Tranter

School of East Asian Studies, University of Sheffield, 6/8 Shearwood Road, Sheffield S10 2TN, UK

Semester 2 Working Days: Monday, Wednesday, Thursday morning, Friday morning

Semester 2 Student Contact Hours: Monday 3:00-4:00pm, Wednesday 11:15am-1:00pm

That is all.