

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 2/WG 2**

**Universal Multiple-Octet Coded Character Set
(UCS)**

**ISO/IEC JTC 1/SC 2/WG 2 NXXXX
2014-02-XX**

| | |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title: | Supplementary Documents for Proposal of Variants for Siddham Script |
| Source: | Deborah Anderson, Lee Collins, Bill Eidson, Andrew Glass, Shoken HARADA (原田聖賢), Taichi KAWABATA (川幡太一), Ken Lunde, Koju MOTOYAMA (元山公寿), Kiyonori NAGASAKI (永崎研宣), Anshuman Pandey, Michel Suignard, Toshiya SUZUKI (鈴木俊哉), Taro YAMAMOTO (山本太郎) |
| References: | See “References” section in the document |
| Status: | Experts’ Contribution |
| Action: | Request for review by ISO/IEC JTC 1/SC 2/WG 2 |

Supplementary Documents for Proposal of Variants for Siddham Script

Abstract

This document summarizes the discussions and conclusions of Siddham ad-hoc meeting held in Tokyo in November, 22nd, 2013. In this meeting, participants concluded that certain criteria must be met for encoding variants. In accordance with the agreement, four of six variants proposed by Japan were concluded that they should be encoded, and Japanese experts will prepare supplementary evidences that meet the criteria concluded in this meeting. This document also provides the supplementary evidences for remaining two variants that satisfy these criteria.

Background

Siddham script is one of Brahmi scripts that was born in 6th or 7th century of India. Japanese historical adaptation of the script evolved to include non-phonetic applications of the script which have not used in the original Brahmic script in India.

As of it, we’ve submitted WG2 N4407R, the proposal to encode variant characters for Siddham script, and it was accepted in WG2 Vilinus meeting and incorporated into DIS ballot text of 4th edition of ISO/IEC 10646.

However, some requests were raised to clarify the criteria for variants to be encoded later. As of it, the Siddham ad-hoc meeting was held in last November, and several criteria are concluded there. In accordance with these criteria, four of six variants proposed in Japan were concluded as appropriate, but for two remaining ideographs, they are left as new ideographs.

Meeting Report

The report of Siddham ad-hoc meeting in Tokyo (L3/13-233) follows. 【

As a neutral party, it was very clear that the right people attended this meeting, and if this meeting didn’t happen, the progress that we made could have taken months or years. And, as usual, having the meeting face-to-face made all the difference in the world. I am thankful that everyone took the time out of their busy schedules to participate.

While the agenda was helpful in guiding the meeting, some items were skipped, mainly because everyone realized what the important points were, such as the criteria for encoding character variants and the status of the six character variants in ISO/IEC 10646 Fourth Edition, and the meeting focused on them.

At the very beginning of the meeting, in order to eliminate any possible confusion, Michel Suignard provided details about the current status of the Siddham script in ISO/IEC 10646. The standard Siddham characters (U+11580 through U+115B5 and U+115B8 through U+115C9) are in Amendment 2 of Third Edition (equivalent to Unicode Version 7.0), and are considered a done deal (frozen). The section marks and character variants are in Fourth Edition, which is undergoing its last technical ballot, and is expected to be finalized during the February 2014 WG2 meeting.

As background, WG2 N4294, which is the original Siddham script proposal (Pandey) states that Siddham is an Indian script that is no longer used in India. We learned during the meeting that the user community is approximately 12 million people, 10 million of which are in Japan. The rest are primarily in China and Korea. Japan also coined digits for Siddham, and Korea has unique Siddham forms.

WG2 N4407R (Japan) proposed six Siddham character variants (U+115E0 through U+115E5), which are reflected in ISO/IEC 10646 Fourth Edition. Professor Motoyama stated that there are no more than 10 character variants (when the agreed-upon criteria, which is effectively the same criteria that Japan used to select these first six character variants, is applied), which means that the current Siddham block is of sufficient size to accommodate them in the future. It was also noted (by Japan) that standardized variation sequences cannot be used for the character variants that are combining forms (U+115E4 and U+115E5).

Siddham ligatures were discussed, including the possibility of encoding the high-frequency ones. There was mutual agreement not to do this, and to instead use font features, such as 'liga' (GSUB). WG2 N4490 (Pandey), which proposed a separate block for Siddham logographic forms, was discussed, but there also was mutual agreement not to do this.

To quote Anshuman, it became clear that Siddham needs to be handled from a Pan-Buddhist perspective. This means that each user community will have their own needs, based on their particular usage of the script. Bill's needs are met by the standard Siddham characters, mainly because character variants are handled via separate font resources. Japan's needs will be met by encoding the six character variants that are in ISO/IEC 10646 Fourth Edition.

Part of the difficulty in handling or interpreting Siddham character variants is that their usage is often based on user will, which amounts to a semantic distinction. Also, Bill pointed out that there are known errors in the historical documents that may produce unique forms, and when passing them down, one must decide whether to propagate, correct, or annotate such errors in the process.

One issue that came up, which also comes up in similar meetings, is how to define plain text. In my experience, one way to think about plain text is to open a PDF that includes Siddham content that is likely to be stylized, copy the text (which is done as plain text), then paste it into a text editor, word processor, or comparable application. If any meaning or information is lost in this process, then the plain text representation is insufficient.

Three very important things came out of this meeting:

- Mutual agreement by all attendees was reached on the following three sets of objective criteria for determining whether a Siddham character variant is suitable for encoding

1. Both forms are semantically distinct in a logographic context.
2. Both forms cannot be algorithmically derived by context.
3. Both forms co-occur in a single source.

• There was also mutual agreement that only general documents should be considered valid sources, and those documents that intentionally use character variants for pedagogical purposes should be excluded.

• These criteria were applied to the six Siddham character variants, and there was mutual agreement by all parties that the following four are encodable:

U+115E0 through U+115E3

Japan feels that they can provide sufficient evidence that the combining character variants, U+115E4 and U+115E5, should be encoded.

• Mutual agreement that a joint document be submitted to WG2 to clarify how Siddham character variants should be handled in terms of encodability, including clarification of the six that are in ISO/IEC 10646 Fourth Edition.

Lee Collins raised the following open technical questions with regard to the handling of the character variants that represent vowels:

- Do we need to define a default or canonical representation of the U vowel sign?
- Which character variant is correct for use with RA, and how do we render the RA followed by non-canonical forms?
- What guidelines will we offer for transliteration software between Romanized Sanskrit and Siddham if we introduce unresolvable ambiguities by separate encoding of U?

】 (L2/13-233, Lunde)

Experts' Response to the Report

In accordance with this conclusion, Japanese experts provide the following evidences of the pending 2 vowel signs. Japanese experts believe that the pending 2 vowel signs are fitting to the criteria in the conclusion of the ad-hoc meeting.

1. Both forms are semantically distinct in a logographic context.
2. Both forms cannot be algorithmically derived by context.
3. Both forms co-occur in a single source.

VOWEL SIGN U (proposed at U+115E4)

The typical distinctive shapes of vowel sign u in the syllable “hu” are shown in Table 1. In following, a curling form with a tail to upper left is called as cloud form (among Japanese Siddham users, described as 雲形). It is used in the code chart of ISO/IEC 10646:2012/Amd.2:2013 code chart as a representative shape of vowel sign u (U+115B2). A sloping form with a tail to lower right is proposed to ISO/IEC 10646:2014. It is called as a warbler form (among Japanese Siddham users, described as 鶯形).



cloud form of the vowel sign u

warbler form of the vowel sign u

Table 1: The Variant Forms of Vowel Sign U

Non-mandala distinction example is shown in Figure 1 that are taken from Usuzoushi by Jougen (薄雙紙 by 成賢 (12th Century), Taisho Vol. 78, Text No. 2495, p. 620-). The hu ligature using the warbler in the first quote is of the mantra for Bhaisajyaguru (藥師如来). The ligatures hu in the second and third quotes are not for Bhaisajyaguru and using the cloud form. It should be noted that the warbler form is used in the matra for Bhaisajyaguru, not the symbol of Bhaisajyaguru in the mandala (in the mandala, Bhaisajyaguru is symbolized by the syllable “bhai”).

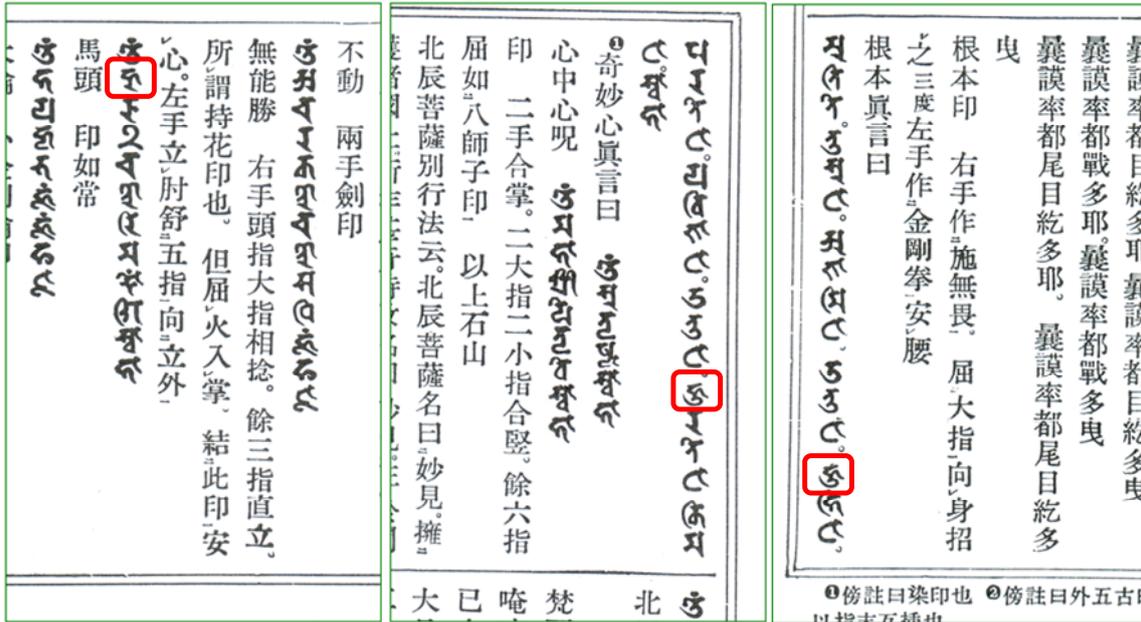


Figure 1: Differentiation of Vowel Sign U in the Mantra for Bhaiṣajyaguru (藥師如來), Usuzoushi, Taisho Vol. 78, Text No. 2495, p. 655 (left) and p.680-681 (mid-right).

Figure 2 is taken from Byakuhoushou by Chouen (白寶抄 by 澄円 (13th Century)), Taisho Vol. 96 (Zuzoubu Vol.10), p. 313). The annotations to the 2 huM ligatures are different; also the surrounding texts describe the different meanings of 2 ligatures. The distinction in this example is not to mean different Bodhisattva or Guru.

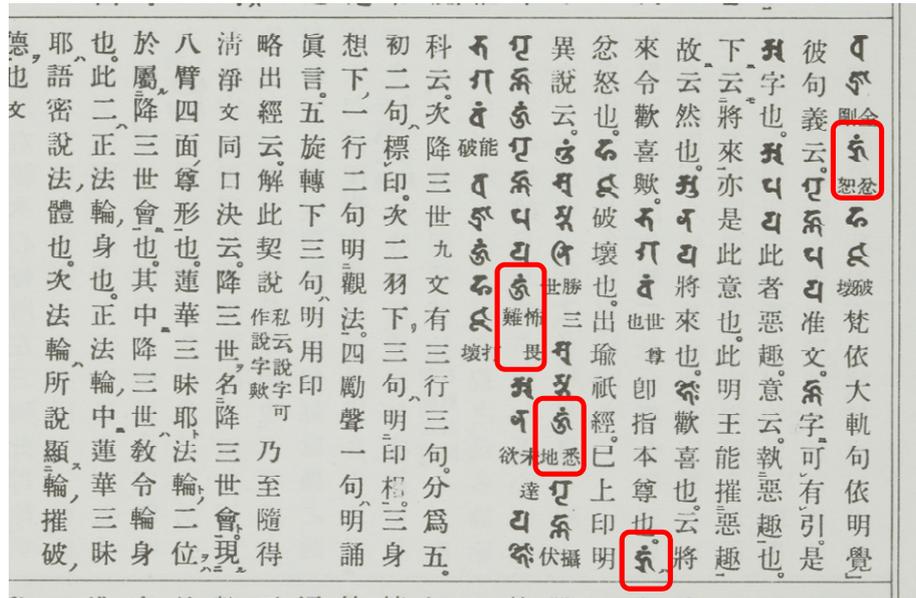


Figure 2: Another Semantic Differentiation of Vowel Sign U in Single Document Byakuhoushou (by Chouen, 13C), Taisho Vol. 96 (Zuzoubu Vol. 10), p. 313

VOWEL SIGN UU (proposed at U+115E5)

The typical distinctive shapes of vowel sign uu in the syllable “huu” are shown in Table 2. As the vowel sign u, a curling form with a tail to upper left with a short branch is called as cloud form in following. Another form looking like a virama sign (in comparison with Siddham virama, vowel sign uu variant is wider and more curling) is called as the warbler form.



cloud form of the vowel sign uu



warbler form of the vowel sign uu

≠

115BF

Table 2: The variant forms of vowel sign uu

Figure 3 (白寶抄 by 澄円 (13th Century)) and Figure 4 (薄雙紙口決 by 頼瑜 (13-14th Century)) show the differentiations of the vowel sign uu in the single document which could not be automatically shaped by the contextual information. Both examples do not seem to be the symbols for specific Bhodhisattva or Guru.

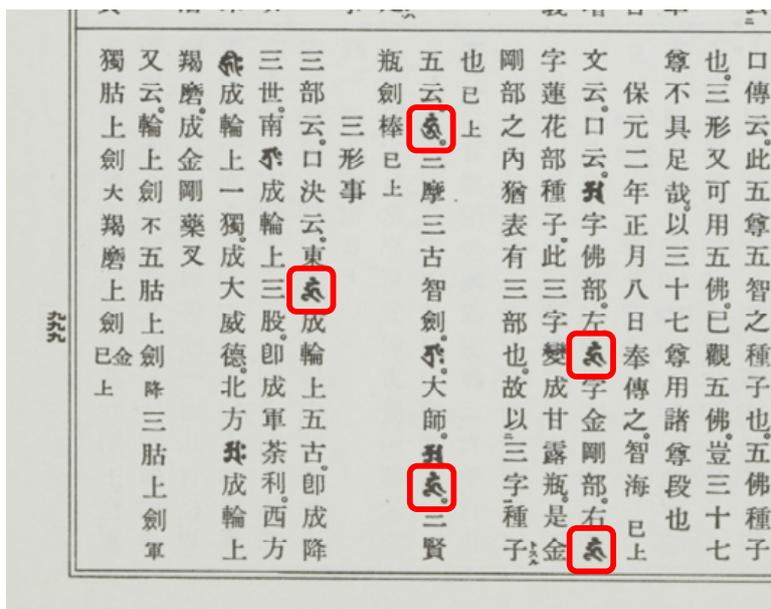


Figure 3: Differentiation of Vowel Sign UU in Single Document
Byakuhoushou (by Chouen, 13C), Taisho Vol. 96 (Zuzoubu Vol. 10), Byakuhoushou, p. 999

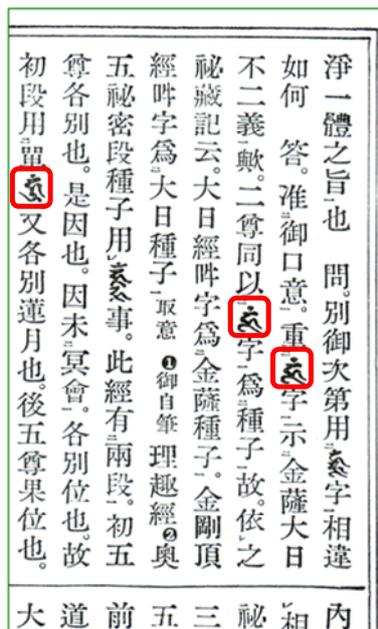


Figure 4: Another Differentiation of Vowel Sign UU in Single Document
Usuzoushi-Kuketsu (by Raiyu, 13-14C), Taisho Vol. 79, Text No. 2535, p. 234

DEFAULT / CANONICAL REPRESENTATION FORMS SHOULD BE DEFINED?

Although Siddham code chart in ISO/IEC 10646:2012/Amd.2:2013 shows the cloud forms for the vowel signs u (U+115B2) and uu (U+115B3), they do not request that the cloud forms should be used for these codepoints. There are some syllables whose u or uu are frequently shaped by the warbler forms, although there is no semantic distinction. Figure 5 shows a sample of a document which never uses the cloud form for the syllable “dhu”.



Figure 5: A Document Which Never Uses the Cloud Form for the Syllable "Dhu". (金記諸会 (1192), manuscript preserved in Toyo Bunko)

Considering that some variant forms do not co-occur with canonical forms in same source, the per-document typeface design should be permitted. Also each school of Buddhism might have their own preferred forms, and they want to make them as default forms in their systems. Therefore a definition of the default or canonical shapes in the international standard of the character encoding would be difficult to be fair and neutral.

LIGATURE OR COMBINING CHARACTER?

WG2 N4486 proposed an option to encode the variant as the ligature for full syllables “hum” and “huum”, to minimize the uncontrolled usage of the variant forms without semantic distinction. Although the examples in documents are for the syllable “hu” or “huu”, but the final consonants for the vowel sign u are varied; hu and hum. Therefore, the ligatures to be coded are not “hum” and “huum”, but “hu” and “huu”. To use them, the extra composition is needed to attach the final consonants. It is difficult to restrict attachable consonants (or we should encode “hu”, “huu”, “hum” and “huum” separately); therefore, the advantage to encode all ligatures separately is not significant.

CHARACTER NAME

WG2 N4490 proposed to include the names of Bhodhisattva into the character names, to minimize the uncontrolled usage of the variant forms without semantic distinction, in spite of the too generic names like “VOWEL SIGN U VARIANT FORM A” proposed in WG2 N4407R. The usage of the variant form should be restricted to the exceptional case that the semantic distinction is needed. Therefore the names should be easy to recognize their exceptional purpose. However, proposed “hu” variant is used for Kundali Vidyaraja in the mandala, but used for Bhaisajyaguru in the mantra text, sometimes it would be difficult to make a fair choice for one referential unique name of Bhodhisattva, Guru, Wisdom Kings or other deities for each variant. Considering that all proposed variants are found as the Biiijaaksara signs in the mandala at least, the names should be in the syntax like “SIDDHAM VOWEL SIGN U ALTERNATE FORM A FOR

BIIJAAKSARA”. In existing names of characters in ISO/IEC 10646, “alternate form” seems to be more popular than “variant”.

Acknowledgements

Authors are thankful to

References

- L3/13-233 (2013-11-22 Siddham Script (梵字) Meeting @ Tokyo, Lunde)
- WG2 N4507 (Comments on the name of the "Siddham" script (L2/12-011R = N4185), Baums & Glass)
- WG2 N4506 (Comments on naming the "Siddham" encoding, Rajan & Sharma)
- WG2 N4490 (A Practical Approach to Encoding Siddham Variants, Pandey)
- WG2 N4486 (Comments on N4407R Proposal to Encode Variants for Siddham Script, Glass)
- WG2 N4468 (Additional Siddham Variants, Pandey)
- WG2 N4467 (Proposal to Encode a Set of Digits for Siddham, Pandey)
- WG2 N4460 (Siddham ad hoc report)
- WG2 N4459 (Draft additional repertoire for ISO/IEC 10646:2014 (4th edition) [see 11580-115FF on pp 23 & 24], Suignard)
- WG2 N4457 (Name changes for Siddham Section marks, Anderson et al.)
- WG2 N4407 (Proposal to Encode Variant Forms for Siddham Script, Kawabata et al.)
- WG2 N4391 (Additional expert feedback on Siddham Section marks, Eidson)
- WG2 N4378 (Additional Information on Siddham Section Marks (N4336), Anderson)
- WG2 N4369 (Feedback on Siddham proposal (WG2 N4294), Eidson)
- WG2 N4361 (Feedback on Siddham proposal (WG2 N4294), Suzuki)
- WG2 N4336 (Proposal to Encode Section Marks for Siddham in ISO/IEC 10646, Pandey)
- WG2 N4294 (Proposal to Encode the Siddham Script in ISO/IEC 10646, Pandey)
- WG2 N4185 (Preliminary Proposal to Encode Siddham in ISO/IEC 10646, Pandey)

Bibliography of Figures

- 『大正新脩大藏經』 [Taisho-Shinshu-Daizokyo], 1924-1934
- 『大正新脩大藏經 圖像部』 [Taisho-Shinshu-Daizokyo Zuzoubu], 1924-1934
- 『金記諸会』, 1192