ISO/IEC JTC1/SC2/WG2 N4490
L2/13-195
2013-10-27

Title: A Practical Approach to Encoding Siddham Variants
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2013-10-27

1 Introduction

The proposal (N4407R L2/13-110) submitted by Taichi Kawabata, et al., sought to encode six characters in the 'Siddham' block (U+11580), which are glyphic variants of basic characters already included in the script. The proposal was approved by WG2 at its meeting in June 2013. The Unicode Technical Committee (UTC), however, did not act upon the proposal at its meeting in July 2013. In consideration of the ongoing discussions regarding the proposal and the importance of the characters to the user community, I wish to offer a practical approach for representing these Siddham variants and other symbols in the Universal Character Set (UCS).

2 Analysis

As stated in N4407R, the rationale for encoding these glyphic variants as separate characters is based upon the view in Japanese Siddham traditions that these forms represent different religious icons and are considered ideographically, therefore semantically, distinct from the normative glyphs. For example, in Shingon teachings the glyphs 𑖃 and 𑖄 represent different aspects of the *bodhisattva* Kṣitigarbha (Jizō), but in terms of the character-encoding model of the UCS, 𑖄 is a glyphic variant of 𑖃 U+11582 SIDDHAM LETTER I.

From a palaeographic perspective, the proposed characters simply are variants of normative glyphs for Siddham characters. Despite culture-specific ideographic interpretations, Siddham is inherently alpha-syllabic in structure and, similar to related Brahmi-based scripts, has its share of historical and regional variants. The distinction between 𑖃 and 𑖄 as different graphical representations of U+11582 SIDDHAM LETTER I is identical to the distinction between অ and अ as variant forms of U+0905 DEVANAGARI LETTER A. Additionally, other variants proposed in N4407R correspond to various historical forms of particular letters. For instance, 𑖅 is proposed as another variant of SIDDHAM LETTER I, but usage of this glyph is not unique to Siddham; it is related to the character 𑰬 BHAIKSUKI LETTER II, which is proposed for encoding in N4489 L2/13-194 as part of the 'Bhaiksuki' block (see also the comparison given there between Bhaiksuki, Sharada, and Siddham).

While the approval of N4407R may be justified according to various UCS principles, the decision did not take into consideration other related characters and encoding issues. In my comments (N4468 L2/13-136) on N4407R I stated that there are several other glyphic variants for Siddham that are used alongside normative forms, in philosophical and pedagogical contexts, which may also be considered ideographically distinct. For example, 𑖌 is an attested variant of 𑖋 SIDDHAM LETTER AA. In the Shingon tradition, the manifestation of the *buddha* Vairocana (Dainichi Nyorai) in the Vajradhātu *maṇḍala* is written not with the normative glyph, but using the variant along with the combining marks *candrabindu* and *visarga* as 𑖌𑖼𑖽. Usage of 𑖌 may be specialized, but it is certainly required for producing seed-syllables and is just as semantically distinct from 𑖋 in an ideological context as are 𑖄 and 𑖅 from 𑖃. For this reason I strongly recommended that no

decision should be taken upon N4407R by either WG2 or UTC until a complete repertoire of variants that possess ideographic values can be established.

Furthermore, the approval of N4407R did not fully take into account the character-glyph model of Brahmi-based scripts in the UCS. According to this model, the proper way to represent these Siddham variants is at the font level, not at the encoding level. Indeed, as Andrew Glass recommends in his comments (N4486 L2/13-189) upon N4407R, the proposed characters should be used only to denote the "supra-phonetic" values of the normative Siddham characters and in all other contexts the display of these glyphs should be managed through fonts. The rationale for this restriction is that the ideographic usage of palaeographic forms, which did not originally possess ideographic values, may lead to requests for encoding palaeographic variants in other scripts as separate characters.

Glass's concern is not simply theoretical. The issue of encoding glyphic variants engages with ongoing discussions regarding the encoding of similar types of graphical forms in other scripts. In the Brahmi family the most obvious case is Devanagari, in which several characters have 'western' and 'eastern' forms, eg. 𑍐 for अ U+0905 DEVANAGARI LETTER A, 𑍇 for ऋ U+090B DEVANAGARI LETTER VOCALIC R, 𑍎 for झ U+091D DEVANAGARI LETTER JHA, 𑍏 for ण U+0923 DEVANAGARI LETTER NNA; as well as digits, ५ for ४ U+096B DEVANAGARI DIGIT FIVE, ८ for ८ U+096E DEVANAGARI DIGIT EIGHT, ९ for ९ U+096F DEVANAGARI DIGIT NINE; and forms of conjuncts, 𑍑 and 𑍒 for *kṣa*; among others. One may find examples in regional Indian esoteric traditions where these variant forms, being regionally-normative, have similar extra-phonetic semantic values, similar to the Japanese ideographic interpretations of Siddham variant forms.

3 Recommendation

Although WG2 approved N4407R, I wish to reiterate the recommendation I made in N4468 and request that UTC *not approve* the proposal at this time. The focus on six variant characters is too narrow, and as described in the original proposal to encode Siddham (N4294 L2/12-234R), there are word ligatures, stroke primitives, pedagogical letters, and other symbols with ideographic values that should be investigated as candidates for encoding in tandem with glyphic variants. Moreover, for organizational and technical purposes, ideographic characters that are palaeographic variants of normative characters of an alpha-syllabic script should not be encoded as part of that script block.

The issue of encoding glyphic variants as independent Siddham characters presents a situation that is unique among the Brahmi-based scripts in the UCS. First, glyphic variants in Siddham have culture-specific semantic values that do not generally inhere in glyphic variants of characters in other Brahmi-based scripts. Second, such distinctive usage is historically attested and continues to be reproduced by users of the script at present. Third, these users request the ability to distinguish between these forms at the character level. Based upon the ongoing discussions, it is clear that there is a requirement to represent both the alpha-syllabic (Indic) and ideographic (Japanese) aspects of Siddham. The most practical method of doing so is to encode the characters in question in a new block to be named 'Siddham Ideographs'.

The 'Siddham Ideographs' block would contain only characters that possess ideographic values. For any such character that is a true glyphic variant, the connection between it and the normative character would be severed and it would be encoded as an independent character and assigned a name that reflects only its ideographic value. For example, 𑖌 could not be named SIDDHAM LETTER I VARIANT FORM A as in N4407R, but must be given a descriptive name such as SIDDHAM IDEOGRAPH KSHITI GARBHA BODHISATTVA FORM-1. For informational purposes, a reference to the normative character in the 'Siddham' block may be added as an alias in the names list. A concept for a 'Siddham Ideographs' block is enclosed. The chart shows the types of characters that may be included and possible naming conventions. It is presented for illustrative purposes only: it is not complete or intended for actual use and the code points are arbitary.

There is an opportunity here to develop a proper encoding for Siddham that will reflect the palaeographic, historical, and cultural aspects of the script and that will meet the needs of user communities across regions and disciplines. If WG2 and UTC agree with my recommendation to create a 'Siddham Ideographs' block, then I welcome input and assistance from the Siddham user community, Japan national body, scholars of Indic palaeography, and other interested groups to begin work on a formal proposal for encoding an initial set of Siddham ideographs.

4 References

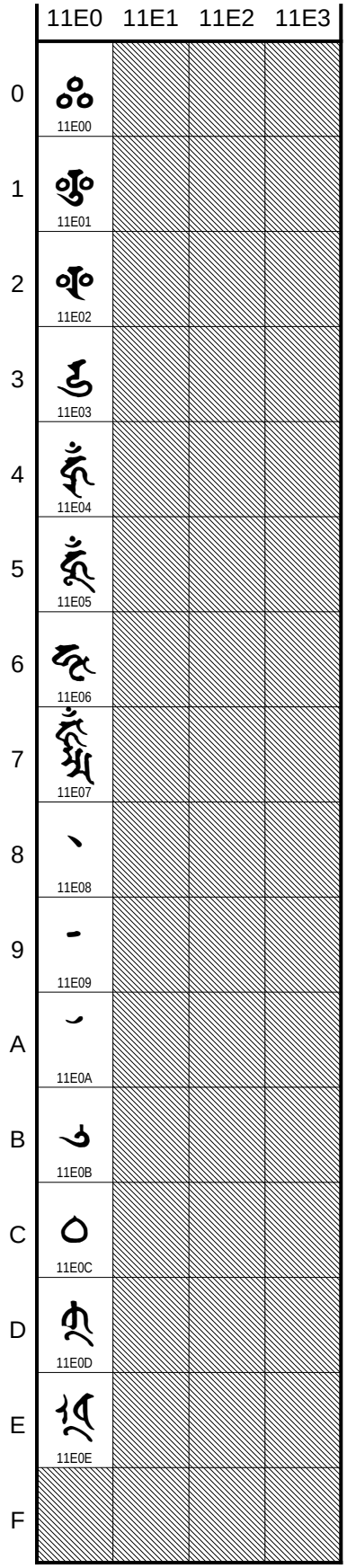
Glass, Andrew. 2013. "Comments on N4409R Proposal to Encode Variants for Siddham Script". N4486 L2/13-189 (October 10, 2013). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4486.pdf

Kawabata, Taichi; Toshiya Suzuki; Kiyonori Nagasaki; Masahiro Shimoda. 2013. "Proposal to Encode Variants for Siddham Script". N4407R L2/13-110 (June 11, 2013). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4407.pdf

Pandey, Anshuman. 2012. "Proposal to Encode the Siddham Script in ISO/IEC 10646". WG2 N4294 L2/12-234R (August 1, 2012). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4294.pdf

———. 2013a. "Additional Siddham Variants". N4468 L2/13-136 (June 15, 2013). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4468.pdf

———. 2013b. "Proposal to Encode the Bhaiksuki Script in ISO/IEC 10646". N4489 L2/13-194 (October 27, 2013). http://std.dkuug.dk/jtc1/sc2/wg2/docs/n4489.pdf



This chart is a concept for a possible block for Siddham ideographs. It is not intended for actual use.

Letter Variants

These characters are glyphic variants of normative glyphs for characters encoded in the Siddham block, but possess ideographic values that are not inherent in the original alpha-syllabic character.

11E00 𑖀 SIDDHAM IDEOGRAPH KSITIGARBHA BODHISATTVA FORM-1
- represents the Zenrin manifestation

→ 11582 𑖂 siddham letter i

11E01 𑖁 SIDDHAM IDEOGRAPH KSITIGARBHA BODHISATTVA FORM-2

→ 11582 𑖂 siddham letter i

11E02 𑖃 SIDDHAM IDEOGRAPH ABC

→ 11583 𑖃 siddham letter ii

11E03 𑖄 SIDDHAM IDEOGRAPH XYZ

→ 11584 𑖄 siddham letter u

Word Ligatures

These ligatures cannot be represented using normative sequences of encoded characters.

11E04 𑖅 SIDDHAM LIGATURE HUM FORM-1
= ligature of <ha, vowel sign u, candrabindu> using a glyphic variant of vowel sign u

11E05 𑖆 SIDDHAM LIGATURE HUM FORM-2
= ligature of <ha, vowel sign u, candrabindu> using a glyphic variant of vowel sign u

11E06 𑖇 SIDDHAM LIGATURE PHAT
= ligature of <pha, tta, virama>

11E07 𑖈 SIDDHAM LIGATURE HAMMAM
- bija of Acala (Budong Wingwang, Fudo Myoo)

= ligature of <ha, vowel sign aa, candrabindu> and <ma, vowel sign aa, ma>

Stroke Primitives

Used pedagogically in the analysis of Siddham letters.

11E08 𑖉 SIDDHAM STROKE MYO-TEN
- life mark

→ 31D4 丶 cjk stroke d

11E09 𑖊 SIDDHAM STROKE A-TEN
- inherent-a mark

11E0A 𑖋 SIDDHAM STROKE GYOGATSU-TEN
- moon-viewing mark

11E0B 𑖌 SIDDHAM STROKE U-TEN
- u mark

11E0C 𑖍 SIDDHAM STROKE EN-TEN
- circle mark

Headless Letters

Used pedagogically in the analysis of Siddham letters.

11E0D 𑖎 SIDDHAM HEADLESS LETTER KA
11E0E 𑖏 SIDDHAM HEADLESS LETTER KHA