**TO:** UTC
**FROM:** Deborah Anderson, Ken Whistler, Rick McGowan, Roozbeh Pournader, Andrew Glass, Laurentiu Iancu, and Lisa Moore
**SUBJECT: Recommendations to UTC #148 August 2016 on Script Proposals**
**DATE:** 30 July 2016

The script ad hoc group met twice in the period from June to July 2016 in order to review proposals. The following represents feedback on proposals that were posted in the Unicode document registry at the time the group met.

**EUROPE**
**1.** *Ogham*
**Document**: L2/16-110 Proposal to define 21 variation sequences for Ogham letters – West

**Discussion:** We reviewed this proposal, which requested 21 variation sequences for representing unusual letterforms, especially for Pictish Ogham inscriptions in Scotland.

Currently, the way to represent the variant shapes is to use either a font with the desired shapes, use PUA, or use Variation Sequences.

With one exception, the examples in L2/16-110 are from secondary sources, specifically drawings of rock inscriptions. Relying on such drawings as the source of standardized variants make verification difficult. Compare the proposed VS for U+1695, where the variation is minor:

 vs. 
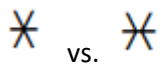
The approach taken in this document is somewhat reminiscent of the handling of Viking Age Nordic runes, where alternate forms were given a different representation (in that case, separate encoding) from the more usual way of handling Germanic and British runes (which relied on fonts to distinguish forms). The approach in L2/16-110 suggests that Pictish is distinct from Irish, and should be handled in a different way.

Would the use of VS help people in the representation of their text? Given the information provided, we feel the distinctions can be handled without encoding more distinctions in plain text.

**Recommendations:** We recommend the UTC review this document, and send comments to the author.

**2.** *Cypro-Minoan*
**Document**: L2/16-179 Revised proposal to encode the Cypro-Minoan script in the SMP (WG2 N4733) – Everson

**Discussion:** We reviewed this document, which is a revision of an earlier document, L2/16-089. The proposal proposes unifying the glyphs from the stages of the writing system, identified as CM1, CM2, and CM3.

A number of questions were raised during the review. Can experts and the proposal author explain why *22 is encoded, given the current assessment of the field, since *22 appears to be a CM3 version of *21? In a similar vein,*19 and *20 appear to be the same character. The proposal should, in our view, reflect

the best understanding of the repertoire, not proposing for separate encoding catalog entries for signs that are generally viewed by expert consensus in the field as being the same. A proposal that includes signs widely viewed as being unified with other characters -- solely to document the history of the decipherment -- is not a good use of an international standard, in our opinion.

**Recommendations:** We recommend the UTC review the document and send the author feedback, along with the comments above.

**AFRICA**
*3. Egyptian hieroglyphs*
**Document:** L2/16-199 A suggestion ... for Ancient Egyptian - Overington

**Discussion:** We reviewed this document, which suggested one new format control character ("EGYPTIAN HIEROGLYPH PARAMETER BASE CHARACTER") followed by either zero or more tag characters and U+E007F CANCEL TAG as a way to get a set of unlimited control characters. We do not feel this is a useful way to create control characters for Egyptian hieroglyphs.

**Recommendations:** We consider this proposal unacceptable, and the UTC should go on record as not wanting it to progress.

*4. Mandombe*
**Documents:** L2/16-077r Proposal for encoding the Mandombe script in the SMP of the UCS – Rovenchak et al.
L2/16-195 Reply to comments raised in L2/16-156 (Mandombe) – Rovenchak

**Discussion:** We reviewed these two documents, the proposal proper and a set of answers to questions posed in the last script ad hoc recommendations (L2/16-156). The proposal is getting closer to a final version.

Comments from the discussion are listed below:

- If every syllable can take a combining mark, the font would need to know the location of the attachment point of each combining mark. Such information could be provided, for example, in an associated Unicode Technical Note (UTN).
- Encode only 7 combining marks, instead of 14 (i.e., don't encode above and below versions of each); a Unicode Technical Note could provide needed specifications for the font designer
- Encode only one acute, instead of four. The attachment point of the acute on the syllable, and whether it is high or low is a placement issue that can be determined by the font designer or specified in a UTN.
- Provide additional evidence in running text of ALTERNATE COLON, PLUS, MINUS, EQUALS, and COMMA ROTATED, and identify the relevant figures in the text (i.e., in §6 Punctuation, mention that an example of EXCLAMATION MARK MIRRORED is provided in figures 3 and 4).
- Give a more detailed explanation on why already encoded arrows couldn't be used for PLUS, MINUS and EQUALS.
- In the answer to the question about gaps in the chart, additional explanation is needed. What does "maintaining character order" refer to? Also, provide more specifics about "character classes."

**Recommendations:** We recommend the UTC review these documents and send comments, including those above, to the author.

*5. Loma*
**Document:** L2/16-201 Update on encoding the Loma script in the SMP – Everson

**Discussion:** We reviewed this update, which provides information on the issues surrounding the encoding of the Loma. No action is requested.

**Recommendations:** The UTC should note this document.

**EAST and CENTRAL ASIA**
*6. Hentaigana*
*a. Hentaigana proposal*
**Document:** L2/16-188 Revised Proposal of Hentaigana (with associated spreadsheet) – Japanese National Body

**Discussion:** We reviewed this revised proposal and spreadsheet, which showed continued progress. We noted that HENTAIGANA LETTER E-1 has been removed, and instead a name alias is being separately proposed (in L2/16-189, see below).

Only a few questions remain on this important proposal. In the cases where a letter has the same sound and the same source, but is not distinguished in the family registration and shows no apparently significant calligraphic difference (such as 257 and 258, below, as well as 9/10, 42/43, 60/61, etc.), a statement should be provided justifying the separate encoding of the character. Such a rationale could, for example, be added as an annotation, explaining why it is distinct.

| Rev# | Proposed Glyph | Revised Character Name | Phonetic Value | Family Reg. Char# | Academic Use Char# | Source ID | Cdpt font |
|---|---|---|---|---|---|---|---|
| 257 | 〈glyph〉 | HENTAIGANA LETTER RO-1 · Derived From 5442 呂 | ろ (U+308D) | 901370 | 430020010 | JMJ-090270 | 1B103 |
| 258 | 〈glyph〉 | HENTAIGANA LETTER RO-2 · Derived From 5442 呂 | ろ (U+308D) | | 430020020 | JMJ-090271 | 1B104 |

**Recommendations:** We recommend the UTC commend the Japanese National Body on its submission, review this document, and relay the comments above to the submitters.

*b. Character name alias for U+1B001*
**Documents:** L2/16-189 Proposal to add a new character name alias to U+1B001 – Japanese National Body

**Discussion:** We reviewed this document, which proposes a new character name alias, HENTAIGANA LETTER E-1, for U+1B001 HIRAGANA LETTER ARCHAIC YE. It also recommends a note be added about the source ("derived from 6C5F") and the glyph be changed (as shown below).

Proposed glyph change: (current) 〈glyph〉 (proposed) 〈glyph〉

**Recommendations:** We recommend the UTC discuss this document as part of the larger Hentaigana discussion. We also recommend an annotation be added to U+1B001, identifying the Hentaigana letter name and source character.

### 7. Sogdian

**Document:** L2/16-158 Preliminary proposal to encode Sogdian in Unicode –Pandey
**Discussion:** We reviewed this preliminary proposal.

A few comments, raised during discussion:
- The encoding of a final-HE seems unusual. Is there no evidence for the occurrence of HE in other positions? Please also point to sources discussing the unusual case.
- Include tables (such as Table 48.2 from World's Writing Systems p. 519) or examples from primers
- Identify the non-Sogdian scripts in figures 3 and 4
- Provide more information on heterograms
- The snail-looking character is described as occurring "in a few sources as an abbreviation for an Aramaic heterogram", but which heterogram is represented? Provide additional analysis.
- Are there known ligatures?
- Provide joining data

**Recommendations:** We recommend the UTC members review this proposal and send the author comments.

### SOUTH and SOUTHEAST ASIA
### 8. Cham
**Document:** L2/16-198 Proposal to encode Western Cham in the SMP – Everson and Cunningham

**Discussion**: We discussed this proposal, which proposes disunification of Cham into Eastern Cham (which is currently in Unicode in the "Cham" block) and Western Cham. Western Cham is used in Cambodia and Vietnam, and Eastern Cham is used in southern Vietnam.

The glyph comparison on page 3 shows that many of the letters are quite different, which strongly suggests disunification would be justified.

**Recommendations:** We recommend the UTC review this document and invite the authors to submit a Western Cham proposal, with illustrations showing the letters (such as from a primer).

### 9. Vatteluttu
**Documents:** L2/16-068 Preliminary proposal to encode Vatteluttu – Pandey
L2/16-118 Feedback on the "Preliminary Proposal to encode Vattezhuththu - L2/16-068" - TVA / T. Udhayachandran

**Discussion:** We reviewed the two documents on the encoding of Vatteluttu, a script used to write Tamil and other languages. The two documents include a preliminary proposal from A. Pandey and the letter from T. Udhayachandran for TVA in response. Based on the comments from TVA, a second proposal with a different model and analysis will be forthcoming.

**Recommendations:** In view of the TVA request, we recommend the UTC not take up technical review of the Vatteluttu, until a TVA proposal is made available and an analysis of the two proposals can be made.

*10. Nandinagari*
**Document:** L2/16-002  Proposal to encode the Nandinagari script – Pandey
L2/16-057  Comments on L2/16-002 Proposal to encode Nandinagari  –  Sharma

**Discussion:** We reviewed these two documents, the proposal proper and comments on it by Shriramana Sharma.

The following were noted:
- Separately encoding digits and headstroke in Nandinagari seems to be well-justified.
- Provide an exhaustive list of all *akhand* ligatures (i.e., those similar to the *kssa* conjunct). If the task is not possible, give a statement characterizing the extent of the author's search, which can might provide a sense of the scope of such ligatures in the script.
- Provide attestation for *upadhmānīya* unification with *jihvāmūlīya*.
- Since Vedic is under-documented, provide as much information as possible on Vedic signs.
- Add 1CF2 VEDIC SIGN ARDHAVISARGA and 1CE9 VEDIC SIGN ANUSVARA ANTARGOMUKHA to §4.5 Script Extensions for representation of *jihvāmūlīya* and *anusvāra* in Nandinagari.
- Figure 8 needs a red box for the other example of *jihvāmūlīya.*
- Highlight and explain the other script in figure 9.
- "Touching forms", discussed in "Conjoined forms", on p. 7ff. will require discussion by the UTC, and impacts the overall encoding model both for Nandinagari and Sinhala (see §1 in L2/16-057 by Sharma).

  The proposal recommends ZWJ + Virama + Consonant to request touching forms of a consonant cluster, which is likewise recommended for Sinhala "touching letters" in "classical and Buddhist texts", but is not discussed in the Sinhala block intro.

  Examples from the proposal show:

  

  However, the general pattern for Indic is:

  A + VIRAMA + ZWJ + B          -->   C1 Conjoining form of A + B

  

  A + ZWJ + VIRAMA+ B          --> A + C2 Conjoining form of B

  

  Clarification about the use of such sequences for controlling touching forms should be taken into consideration for Nandinagari and Sinhala.

- Minor typos:
  - In §3.7 Conjuncts (p. 6), the glyph for U+200D should be ZWJ (not ZWNJ)
  - On p. 8, the sequences for *yya* should have YA, instead of RA

Other comments, based on §2 from Sharma's document L2/16-057
- Take into consideration the comment in §2.1 about shape of LA
- As noted in §2.2, clarify the source of figs. 17-26  (pp. 32-41)
- Explain or fix glyph for *svarita* in §3.15 (p. 11), as it appears to be at variance with figure 12 (p. 28) (Sharma's comment  §2.3)
- Fix glyph for LA (from current glyph for NA) on page 7 (under "Subjoined" in <GA, VIRAMA, LA>) (Sharma's comment §2.3).

**Recommendations:** We recommend the UTC review this document and discuss the "touching letter" issue for Nandinagari and Sinhala.

**INDONESIA**
*11. Buginese*
***The following three documents request the addition of additional characters to Buginese for various orthographies. The full set of proposed characters appears in (d) Buginese Extensions.***

*a. Sumbawa*
**Document:** L2/16-096 Representing Sumbawa in Unicode – Pandey

**Discussion:** We reviewed this document, which recommends adding characters to Buginese in order to represent the Sumbawa orthography, used to write the Sumbawa language, a Malayo-Polynesian language spoken in Indonesia.

Currently, the Buginese block in the BMP has only two open slots remaining. However, it is recognized that adding additional characters in the SMP and hence splitting Buginese between the two planes can cause problems. Ideally, a solution should be arrived at that avoids the situation that arose with Khamti, i.e., where one community uses a specific glyph appearance, and doesn't want the other glyph forms to appear.

In order to make a decision if the letters should be separately encoded or handled as variants in fonts, more information is needed.

Comments raised during discussion included:
- Consider the set of archetypes, pulling together the variants within the letterforms that appear in books and manuscripts, and then decide which letters are necessary.
- Good candidates appear to be: BA, ZA, SYA, FA, QA and DDA.
- Draw on evidence beside what appears in fonts.
- Provide more evidence for the 5 alternate characters on the top of page 5 (eastern and western ja, eastern and western a, and western ra, also listed on page 6, item 5).
- Show different forms of the killer, provide rationale if more than one is proposed (cf. script ad hoc recommendations on the Buginese viramas from May 2016, p. 12, L2/1-156)
- Minor typo:
  On page 2, the following should be /u/:

**Recommendations:** We recommend the UTC review the document, and forward comments to the author, including those above.

*b. Bima*
**Document:** L2/16-119 Representing Bima in Unicode –Pandey

**Discussion:** We reviewed this document, which recommends adding additional characters to Buginese to represent the orthography of the Bima language, a Malayo-Polynesian language spoken in Indonesia.

The following comments were made during discussion:
- Of the consonants proposed for separate encoding, it was not clear the full list should be separately encoded. GA, for example, shows only a small difference in its glyph, cf.:
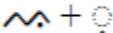
|  | Buginese | Bima |
|----|----|----|
| GA | 〜 | 〜 |

  One letter is quite different (and not in the Buginese block):

  YA        〜〜        ×

  FA, NTA and GEMINATION MARK also appear to be candidates, since they are not in Buginese.

  In order to make a strong case, provide evidence beyond web sources or fonts. What is the range of forms in these sources (realizing that writing on palm leaves may impact glyph representation)?

- Show the vowel silencer in running text and describe its use.  The $\overset{\wedge}{\circ}$ glyph doesn't match the shape in figure 2 (▨).

- Minor typo:
  On page 2, the following should be /u/:

  /i/        〜        〜 + ◌

**Recommendations:** We recommend the UTC review the document, and forward comments to the author, including those above.

*c. Lota Ende*
**Document:** L2/16-076 Representing Lota Ende in Unicode – Pandey

**Discussion:** We reviewed this document which, like Bima and Sumbawa, suggests the addition of several characters to the Buginese script is needed to write the Ende language of Indonesia.

It is clear from the figures that Lota Ende is using the Buginese script with extra letters. A next step would be to provide a document which provides an overall model, showing which letters are unified with which, what alternate forms there are, etc.  (See comments under [d] Buginese Extensions, below.)

Other comments:

- The distinction for NGA is not clear (see below), but CA and SA may be eligible.

|  | Buginese | Ende |
|---|---|---|
| NGA | 人 | 人 |

- The set of characters not found in Buginese and listed on the bottom of page 4 appear to be good candidates.
- Provide an example of the vowel silencer and point it out in the example.
- The differences shown in the chart between Ende and Buginese in figure 8 may be due to pen choice, and not necessarily be considered strong support for the separate encoding of a character.
- Minor typo:
  On page 2, the following should be /u/:

  /i/    ᨆ    ᨆ + ◌

**Recommendations:** We recommend the UTC review the document, and forward comments to the author, including those above.

### d. Buginese Extensions
**Document:** L2/16-159 Preliminary proposal to encode 'Buginese Extensions' in Unicode – Pandey

**Discussion:** We reviewed this document, which pulled together characters proposed in the documents on Sumbawa, Bima, and Lota Ende, and the viramas in L2/16-075.

While legitimate extra letters do appear to be needed for the orthographies using the Buginese script, the proposed set seems to be overly disunified. A systematic analysis of the entire set of Buginese extensions in a single document is needed, one that provides enough evidence to determine which additional letters are truly distinct, and which are variant forms of characters already encoded for Buginese.

**Recommendations:** We recommend the UTC review the document, and forward comments to the author, including those above.

### 12. Lampung
**Document:** L2/16-073 Preliminary proposal to encode the Lampung script – Pandey

**Discussion:** We reviewed this preliminary proposal for the Lampung script used in Sumatra, Indonesia, and used to write the Lampung languages.

Lampung shows very close affinity to Rejang. The historical evidence provided for Lampung varies. For example, figure 6 mentions that /i/ is indicated by a ˚ above the consonant, but the proposed glyph is

◌̆ . Similarly, figure 6 describes /u/ as represented by a ˌ below the consonant, but the proposed glyph is ◌̠ .

In order to make an informed decision on whether Lampung should be separately encoded, evidence needs to be provided showing how the Lampung and Rejang have split in systematic ways.

Other recommendations:
- Provide examples of modern evidence besides signage.
- In a comparison chart, include palm leaf examples (for example, figure 3 shows very Rejang-like angular shapes of the letters)
- Provide the correct transliteration in the captions for figures 17 and 18.

**Recommendations:** We recommend the UTC review the document, and forward comments to the author, including those above.

### 13. Kerinici
**Document:** L2/16-074 Preliminary proposal to encode the Kerinci script – Pandey

**Discussion:** We discussed this preliminary proposal for Kerinci, a script used in Sumatra, Indonesia. The script is clearly differentiated from Rejang, as shown in Table 1, and hence appears well-justified for encoding.

The following are comments raised during discussion:
- Transcribe fully the inscription in figure 4
- The example in figure 5 shows a Latin digit. Are they typically used?
- What are the circles in figure 2 and the middle image in figure 5?
- When discussing the alternate forms, provide evidence

**Recommendations:** We recommend the UTC members review the document, and send the author comments.

### SYMBOLS
### 14. Half star symbols
**Document:** L2/16-186  Proposal to encode four half star symbols – West

**Discussion**: We reviewed this proposal, which proposes four half star symbols, such as those found in ratings. The topic was raised recently on the Unicode email list, but, as noted in the proposal, had been discussed earlier, in 2012.

In our view, the proposed names are suitable. The two symbols which have an outline (STAR WITH LEFT/RIGHT HALF BLACK) are consistent with the names of similar characters, such as:

25D0  ◑  CIRCLE WITH LEFT HALF BLACK

The location, in the Geometric Shapes Extended block, is also acceptable.

**Recommendations**: We recommend the UTC accept the following four half-star characters, after discussion:
1F7D9 LEFT HALF BLACK STAR
1F7DA RIGHT HALF BLACK STAR
1F7DB STAR WITH LEFT HALF BLACK

1F7DC STAR WITH RIGHT HALF BLACK

### 15. Go Notation Symbols
[L2/16-185](#) Proposal to encode symbols for Go game notation (WG2 N4719R) – West

**Discussion:** We reviewed this proposal, which is a revised version of L2/16-108. The latest proposal incorporates comments from the May 2016 UTC discussion, i.e., the problematical character COMBINING ENCLOSING NEGATIVE CIRCLE was removed, and the characters names made more consistent with others already in Unicode.  The proposed names and location (in the Geometric Shapes Extended block) are deemed to be fine.

**Recommendations:** We recommend the UTC accept the following four proposed characters, after discussion:
1F7D5 CIRCLED TRIANGLE
1F7D6 NEGATIVE CIRCLED TRIANGLE
1F7D7 CIRCLED SQUARE
1F7D8 NEGATIVE CIRCLED SQUARE
We also suggest annotations to similar forms be added (i.e., U+238A) and mention of the Humanist movement symbol in the annotation of U+1F7D5.

### 16. Astrological Symbols
#### a. Additional Symbols for Astrology
**Document:** [L2/16-080r](#) Additional Symbols for Astrology (revised) – Faulks

**Discussion**: We reviewed this proposal, which was revised in light of comments from the May 2016 script ad hoc recommendations ([L2/16-156](#)). A few characters from the earlier version were removed, more documentation was provided (identified by comments in green or examples with a green bar), and one new character was added, TRUE LIGHT MOON ARTA. Printed sources for PHOLUS and NESSUS have now been provided, as are examples of the new TRUE LIGHT MOON ARTA.

The proposed seven characters supplement the twelve new astrological symbols on the PDAM1 ballot (for Pluto and the Uranian planets).

We note that HYGIEA is spelled differently from  U+1F54F BOWL OF HYGIEIA, but HYGIEA appears to be the usual spelling for the asteroid, though the author comments in §4 that this asteroid is commonly misspelled HYGEIA.

The proposed names and locations of the characters in the Miscellaneous Symbols and Arrows block appear to be acceptable. The author originally suggested moving the entire set of astrological characters to a block in the SMP, but this was not discussed by the ad hoc, and the PDAM1 ballot already contains several new astrological symbols in the Miscellaneous Symbols and Arrows block.

**Recommendations**: We recommend the UTC review this proposal and accept the following nine characters:
2BE8 TRANSPLUTO
2BE9 PROSERPINA
2BF0 ASTRAEA
2BF1 HYGIEA

2BF2 PHOLUS
2BF3 NESSUS
2BF4 WHITE MOON SELENA
2BF5 BLACK DIAMOND ON CROSS
2BF6 TRUE LIGHT MOON ARTA

### b. Eris and Sedna Symbols
**Document:** L2/16-173 Eris and Sedna Symbols – Faulks

**Discussion**: We reviewed this proposal, which requests three symbols for these trans-Neptunian objects: two for Eris and one for Sedna.  The characters all are documented, and the names are appropriate.

**Recommendations**: We recommend the UTC review the proposal and, after discussion, accept the following three characters:
2BD8 ERIS FORM ONE
2BD9 ERIS FORM TWO
2BDA SEDNA

Since Sedna is derived from a ligature of two Unified Canadian Aboriginal Syllabics, we recommend a cross-reference to the relevant characters be added.

### SHORTHAND SYSTEMS
### 17. Pitman Shorthand
**Document**:  L2/15-116 Encoding Pitman Shorthand Scripts – Ramachandran

**Discussion:** We reviewed this document, which has gone through several revisions. The latest version shows continues to show progress.

The following comments were raised during discussion:
- Separately encode the full vowels
- Reconceive the repertoire of halved and doubled consonants as sequences of the basic letters plus either a halving combining mark or a doubling combining mark; such combining marks would act as diacritic elements modifying the stroke in front of them
- Remove the 3 position controls
- Explain more fully the prefix and suffix characters ("{0}" and "{4}")
- The use of joiners and non-joiners shown in the discussion on the orientation of diphthongs on page 27 doesn't match use in Duployan and is, in our opinion, unnecessary and confusing, unless there are minimal pairs that need to be distinguished.

**Recommendations:** We recommend the UTC review this proposal and send any comments, including the comments above, to the proposal author.

### COMBINING MARKS
### 18. Combining Diacritical Marks
**Document**:  L2/16-178   Representative glyph and annotation additions for U+033B – Evans and Keating

**Discussion:**  We reviewed this document, which requested the glyph for U+033B be changed, and two annotations added.

COMBINING SQUARE BELOW was in Unicode 1.0, with a "squarish" glyph. The Unicode 9.0 glyph also appears to be square:

Unicode 1.0 glyph:     Current (9.0) glyph: 033B

A quick survey of current fonts shows that the glyphs for U+033B can vary.

As noted in the document, this character is part of the set U+033A and U+032A which demonstrates that U+033B is built on the same shape, but closed up with an additional horizontal. (U+0346 is also a member of the set.)

| 033A | ○̺ | COMBINING INVERTED BRIDGE BELOW |
| 032A | ○̪ | COMBINING BRIDGE BELOW |
| 0346 | ○̆ | COMBINING BRIDGE ABOVE |

The use of the character and its square shape are documented in the proposal (citing the following from the *Handbook of the IPA*):

| | Subscript bridge | Dental | n̪ | d̪ | 408 | 032A | E22C |
|---|---|---|---|---|---|---|---|
| | Inverted subscript bridge | Apical | n̺ | d̺ | 409 | 033A | 23FD |
| | Subscript square | Laminal | n̻ | d̻ | 410 | 033B | 23FE |

**Recommendations:** We recommend the UTC review this document, issue a glyph erratum, change the glyph for U+033B in the code chart, and add two annotations, "IPA: laminal", and a note that the rectangular shape is based on the set U+033A, U+032A, and U+0346. An annotation "IPA: apical" for U+033A INVERTED BRIDGE BELOW should also be added.

**The following documents were not discussed by the Script Ad Hoc:**
- L2/16-174 Extra Aspect Symbols for Astrology
- L2/16-209 Proposal to Encode an Abbreviation Sign for Gurmukhi
- L2/16-210 A system of control characters for Ancient Egyptian hieroglyphic text
  (and any documents beyond L2/16-210)