

To: UTC

From: Debbie Anderson, SEI, UC Berkeley

Subject: Outstanding Western Cham Issues (based on 25 May 2020 version of proposal by Martin Hosken)

Date: 9 July 2020

1. Need to decide where to put the eight lunar symbols (Doug Ewell PRI feedback, [L2/20-104](#)), These are currently located in Arabic Mathematical Alphabetic Symbols block. [Ad Hoc suggestions from the April recommendations](#) include moving them to the Western Cham block, the Arabic Presentation Forms blocks, or in a new block.

2. Comments from [L2/20-122](#) still not addressed:

Chart: Request by Script Ad Hoc to provide a chart with the characters. What is the chart in figures 49 and 50 from? These charts were in the Jan version of the proposal.

U+1E263 SIGN TANA TAMAT AYET: Script Ad Hoc requested this character (described by Script Ad Hoc as “at the boundary between Arabic and Western Cham”) be moved to Supplemental Punctuation block. It is still in the Western Cham block. The author asks the UTC to justify its requirement that character be moved and compare with other scripts that share characters between scripts not including Latin.

U+1E264 SIGN TANA TAMAT TAKUE: Script Ad Hoc asked if there are examples of the character beside Arabic text? No new figures are provided. Martin responded via email to Debbie A on April 30 (after UTC): “There are clearly no such examples, so no action required.” Mention in the proposal that there are no other examples known.

U+061D ARABIC END OF TEXT: The Script Ad Hoc stated this character appeared to be a sequence of U+1E263 SIGN TANA TAMAT AYET, and recommended it be removed.

The proposal states, “[ARABIC END OF TEXT] is distinct from a sequence of U+1E263 (SIGN TANA TAMAT AYET) as can be seen in Figure 51”

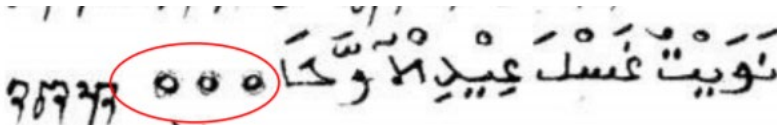
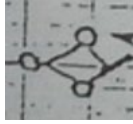


Fig 51. Examples of sequences of U+1E263 (SIGN TANA TAMAT AYET). Notice the space after the left most one on line 2.

U+1E25F TRIPLE DANDA: Script Ad Hoc regards the triple danda in figure 20 as a sequence of double and single dandas, and hence asked it to be removed. New proposal has same text as the March 25 2020 version and retains the character. The current proposal states: “This is a single concept to users rather than some arbitrary sequence of double and single dandas that a user might want to type, making it effectively impossible to search for and select the triple danda.” No new evidence to support it, though Martin reports the users really want this character.

Figure 23: Question from the Script Ad Hoc on the character (below). No information provided. Since it appears in text, it would be useful to find out its meaning/use. Can Martin ask the community?



Figures 24-25: Question about the context of the figures, stamped by the Chief Imam. Was he asked to write down the vowels and consonants by the proposers? Is it a chart he uses for instruction? What is the date?

(Figure 24 is labelled “Consonants with imprimatur from the Chief Imam” but the a handful of the last set of characters are vowels, not consonants.)

Figure 19 shows an error mark represented by U+0353, but this is not discussed in the prose section of the proposal.

Question about an example showing U+1E21B PPA, but no clear example is provided. Martin was to ask the community.

3. Question from Martin (page 9 of new proposal):

In review, it is noted that 1E233 and 1E236 are both right medials. The Universal Shaping Engine has the unfortunate constraint of only allowing one medial in each position. The author is open to suggestions as to how to resolve this problem. Either the USE has to change to support multiple medials in a position, or one of the right medials has to be arbitrarily marked as above just to conform to a technological constraint. It should be noted that allowing multiple right medials would not cause any ambiguity in the erroneous situation that a user enters multiple right medials where only one is expected, since right medials are assumed to be spacing and not attach to the base. But, in discussing this, we are way beyond the scope of this document.

4. Other comments (mentioned in Script Ad Hoc Recommendations [L2/20-105](#))

a. Printed evidence

The proposal states: “It should also be noted that there is very little print material in Western Cham. All the core samples are taken from handwritten books and documents.” Other publications have appeared in print, some of which appear later than those cited in the proposal’s bibliography. These publications are cited and listed in [L2/20-018](#) as well as in the 2016 proposal, [L2/16-198](#). While the proposers may not agree with the group that published that material, it is still printed text. As a result, the proposal does not appear to reflect the printed evidence for the script as a whole, but rather one particular community’s viewpoint.

b. Community Division

“The proposal in N4734 has been closely reviewed by script experts and community leaders among the Western Cham in Cambodia and this proposal arises from their considerations.”

The response to Pérez Pereiro’s document [L2/20-018](#) is as follows (from Martin Hosken’s [L2/20-062](#))

The difficulty in this case is that the Imam San community considers the analysis in L2/20-018 faulty and so the addition of the proposed characters as redundant at this time. There is interest from the Imam San community in engaging with the Cham Language Advisory Committee to see what might be done to address real issues in the script and help it to move forward.

Mention that there is a difference among the users. What exactly are the key technical issues (deemed “faulty analysis”) between the two communities? As noted in the Script Ad Hoc recommendations [L2/20-105](#), “We encourage the two groups represented by the proponents of the proposal and L2/20-018 to collaborate.”

6. New text on logical vs visual model, as requested by Script Ad Hoc [L2/20-105](#) (new text added to proposal on page 8; text copied below):

The encoding model follows that of Eastern Cham and Khmer, in being Indic with vowels following base characters, as described in the cluster patterns. An alternative would be to use a visual encoding order where the medial ra would be encoded before the base consonant. This makes no logical sense in that none of the confusion arguments used for visual ordering apply here. Could a visual ordering be used? Yes. But it would confuse everyone and have to answer the question why differ from Eastern Cham and Khmer when there is no need?

7. Other points (from Norbert L)

– There are a few character name inconsistencies, likely victims of repeated renaming, and I’m not sure that’s the complete list:

- U+1E260 is spelled SIGN TANA TAMA PHON twice on page 2, SIGN TANA TAMAT PHON on pages 2, 7, 11, 47.
 - U+1E262 is spelled SIGN TANA CAPEDAEM IU on page 2, SIGN TANA CHAPEDAEM IU on pages 2, 7, 11, 18 (twice), SIGN TAHA CHAPEDAEM IU on page 19.
 - U+1E24D is named CONSONANT SIGN FINAL H twice on pages 3, 8, FINAL H on pages 4, 5, LETTER FINAL H on pages 6, 11, CONSONANT LETTER FINAL H on page 26.
 - U+1E25A is named SIGN MOON TEN on page 3, DIGIT MOON TEN on pages 7, 11.
 - U+1EEF8 through U+1EEFF are named LUNAR 1 DOT etc. on pages 3, 4, ONE DOT LUNAR MONTH etc. on page 7, LUNAR 1 DOTS on page 15.
 - U+1E201 is named A on page 4, LETTER I on page 4, LETTER E on pages 6, 11.
- Do the ligatures BRAY, NAN, SHALOM ever accept additional marks? The regular expression on page 8 seems to say they don’t. If that’s the case, the Indic syllabic category Other may be more appropriate than Consonant_Placeholder.
- Shouldn’t SHALOM be SALAM, as the expanded character sequence would indicate? Or is the pronunciation really closer to “shalom”?

– The sequences 𑄓𑄓𑄓𑄓𑄓𑄓𑄓𑄓 in the bibliography seem unlikely

Additional comment from Debbie Anderson

There are two figure 32’s and 33’s

Comments from Patrick Chew

See separate document.