Universal Multiple-Octet Coded Character Set
International Organization for Standardization
Organisation Internationale de Normalisation

Doc Type:	**Working Group Document**
Title:	**Response to Western Cham Script Adhoc Comments L2/20-046**
Source:	**Martin Hosken**
Status:	**Individual contribution**
Action:	**For consideration by UTC and ISO**
Date:	**2020-01-25**

Introduction. Not all responses to adhoc comments are best handled in the proposal itself. This document addresses those comments which are not handled by the proposal. It is noticeable from the comments that this proposal has received a much deeper reading than any other proposals the proposer has submitted in the past. This is a good thing, resulting in many helpful comments. But it must be born in mind that the proposer only has limited time to polish proposals. The key aim of a proposal is to see characters encoded and not necessarily to write the perfect paper.

Justify the encoding model. There are two parts to this question. The first is that the proposer is expected to justify their encoding model against some arbitrary encoding suggestion, hinted at in the adhoc comments. If the proposer of this alternative model feels strongly that their suggestion is in the best interests of the community, the encoding and the standard, then the author would welcome them making a formal counter proposal and would engage with that. On that basis, all discussion of such a final character marker has been removed from the proposal.

The second question of justifying why a logical order is required seems counter to all other proposals where a visual order has to be justified because a logical order is presumed. It is not the duty of a proposal writer to justify all the presuppositional bases of the Unicode standard and its principles for every new character and script. Again, if the logical order is considered wrong, then a counter proposal would be welcomed. As to justification, the statement that Western Cham follows the same encoding model to Eastern Cham is considered sufficient.

Provide a chart. The charts provided have been stamped with the authority of the community leader. They are about as authoritative as one can get. They are certainly more authoritative than something published in a book that hasn't been checked by the community. Figures 49 and 50 have been added for those readers who must have a page number on a sample document. No attempt has been made to find the particular book references.

Indic Positional Categories. The author sees no incompatibility between the rendering of the font and the Indic positional categories proposed.

- 1E234 sign ra is not bottom and left, especially when compared with other bottom and left characters. It corresponds most to U+103C (which is not in the positional categories files in Unicode 12.0.0) or U+1A55.

- U+1E235 sign la and U+1E232 sign u are category bottom since when they are combined into a cluster they take no space. But their position is appropriate as presented in the font. Compare with U+0BC1, U+0E3A. But it is acknowledged that particularly the shape of U+1E232 in the chart could be interpreted as being in right position.

- If one looks at the kinds of characters that have indic position of top-and-right one would not categorise anything in this script as top-and-right.

USE support. The adhoc comments that the ordering given is incompatible with the USE. Here are the categories in encoding order:

Codes	Category
1E200-1E228	B
1E234	Mpre
1E235	Mbelow
1E233	Mpost
1E236	Mpost
1E22F-1E230	Vpre
1E232	Vbelow
1E229-1E22C 1E231	Vabove
1E22E	Vabove
1E23D-1E24D	B

From the author's understanding of the USE, this sequence should work with the USE. The author would be interested in understanding what lead to the confident assertion that this script will not work in the USE.

Figures. The debated shape in Figure 23 has been analysed as a non-textual decoration that they do not want to be encoded.

Moving punctuation characters. One school of thought considers that all punctuation characters that might be used in other scripts should go into a common block of punctuation. It is noticeable that everything in the supplemental punctuation block is there because it shares between at least one script and Latin script. Here there is no sharing with Latin script. The proposal follows the school of thought that one follows the script and keeps as much of the script in one block as one can. This is particularly true for non-Roman scripts. Therefore the proposer asks the UTC to justify its requirement that the character be moved and compare with other scripts that share characters between scripts not including Latin.

Tkaj Ka. There has been some confusion on the author's part over this character. But it is a final. This begs the question why it is therefore not listed among the finals. First it is confusible with U+1E24D (SIGN FINAL H), which encouraged moving it away from that spot. The rest of the finals list is encoded to parallel the list in Eastern Cham. While none of these are very strong reasons for not moving it, neither is it a problem where it is. If it has to move, then the recommendation is that U+1E24D be change general category from Mc to Lo with a requisite name change to LETTER FINAL H and Tkaj ka move to U+1E24E.

Conclusion. It is strongly hoped that this proposal is the final version. The language community is very small and has been worn down by the incessant request for yet more examples and justification. There is a reason Socrates was executed. Looking forward, there are signs that the two schools of thought on how Western Cham should be written using the script may resolve their differences. On this basis, there may well be a proposal to add what the discussions decide is missing. But without an initial encoding in Unicode, such discussions may not happen. Communities need to be able work with what they have if they are to be able develop. On this basis, this proposal is proffered as sufficient for the needs of the Western Cham as they understand their script at the moment.