Feedback from Tamil Experts on Tamil Character Names and Annotations

of Symbols and Fractions

I forwarded the document http://www.unicode.org/L2/L2013/13210-script-rec.pdf to Tamil Experts who serve in the Tamil Nadu Govt. specialists' committee and Tamil scholars. It was done on Friday, November 1, 2013 asking for a review of Tamil section on L2/13-210 and their expert feedback from Tamil Nadu authorities on naming Unicode Characters for Tamil in SMP block. As the time is short, I also spoke with many of the officials and specialists of Tamil script on November 3, 2013. Here is a summary of the Tamil Experts' feedback. They are willing to answer any question that UTC and WG2 have regarding the Tamil Nadu Experts' committee recommendations. They include (a) Dr. V. C. Kulandaiswamy, Chairman-Founder, Tamil Virtual University, Tamil IT Department, Govt. of Tamil Nadu (b) Dr. S. Rasu, Professor Emeritus, Dept of Tamil and Epigraphy, Tamil University, Tanjore (c) Prof. K. Nachimuthu, Chairman, Tamil Department, Central University, Tiruvarur, Tamil Nadu.

A. Using Unvoiced Consonants (k,c,t,p) instead of Voiced Consonants (g,j,d,b) in Tamil character names:

"While we agree that it is not clear what Tamil transcription standard is being referred to, the Tamil Virtual Academy's recommendation seems to be based on dictionary forms, such as those in the University of Madras Tamil Lexicon, where initial c is the preferred spelling (instead of initial s), initial p (instead of initial b), medial k (instead of medial g), medial t (instead of medial d), and the long vowels spelled out. By relying on the entries in a standardized reference (such as a dictionary), the spelling in the Unicode Standard will be as consistent as possible. "

The Tamil Nadu experts appreciated very much Deborah Anderson et al., recommendation to use Tamil based transcription method to be used in **Tamil Character names and annotations**. Yes, they recommend (1) using –p-, instead of –b- (2) using –c- instead of –j- (3) using –k- instead of –g- (4) using –t- instead of –d- etc., and also using long vowels spelled out. They not only looked at Madras Lexicon methods, but also the Government practice of writing in Land registration documents and Court records. Overall, not just Lexicon, but the Tamil Nadu Govt. practice is used as the guide to send the Tamil Nadu Govt. recommendation submitted earlier (L2/13-161, by Tamil Virtual Academy, Govt. of Tamil Nadu).

Hence, they specifically mentioned (1)Indian rupee's annotation with "ruupaay" for U+0BF9. (2) "pillaiyaar cuzhi" as annotation for U+0BF3 (3) "naalatu" as annotation for U+0BDF . (4) "naazhi/pati" for U+0BF3 annotation (5) "varutam" for U+0BF5 (6) "etuppu" for U+0BF7 (7) "merpati" for U+0BF8 etc. etc.,

Tamil Nadu Experts' feedback is the appreciation of the recommendation to use consistent Tamil transcription in Unicode standard for symbols and fractions.
"the spelling in the Unicode Standard will be as consistent as possible. " – This principle should be done & we can produce a revised spelling table for to be used in L2/13-047 mainly in using unvoiced consonants, -p-, -c-, -k-, -t-, -d- as Tamil transcription recommendation by Tamil Nadu.

(2) Using KECAM or KESAM as Character name:

"For the representation of certain liquids and nasals, such as r̲, l̤, ḷ, ṇ and ṉ, the spelling RR, LLL, LL, NN and NNN is the convention already present in the names list. The ASCII-based spelling of these letters (r̲, l̤, ḷ, ṇ, and ṉ) when they appear geminated would be, we agree with Sharma, rather awkward. As noted in **L2/13-175,** piḷḷai would be spelled PILLLLAI.

For GEJAM / KESAM, see **§8** below.

Recommendation: We recommend the UTC adopt the current spelling described above (except for the liquids and nasals mentioned), and support a ballot comment for PDAM1 reflecting this. For the cited liquids and nasals, we recommend the UTC invite the TVA to prepare acceptable suggestions, particularly on how to deal with these letters when geminated. We also recommend the changes to the annotations be directed to the Editorial Committee for review. "

As Deborah Anderson et al., said "the spelling in the Unicode Standard will be as consistent as possible. " This can be done by using Unvoiced consonants in English alphabet. And, we can work with Editorial Committee to achieve Tamil character names and annotations as consistent as possible based on Tamil Nadu recommendations.

"8. Name Change to One Character
L2/13-176 Request to change one character name from N4430 L2/13-047 (WG2 N4478) – Sharma
change one character name in N4430 L2/13-047)
Discussion: We reviewed this proposal, which requests a name change from GEJAM to GAJAM.

Recommendation: We recommend the UTC change the name to KEJAM, which is one of the spellings found in the University of Madras Tamil Lexicon. We suggest the UTC support a ballot comment for the PDAM1 to this effect. Because the word has alternate transliterations (as mentioned in **L2/13-176**), the UTC may consider entertaining a request to add the alias KAJAM."

Tamil Nadu Experts appreciate changing the word-initial g to k as is done in Govt. documents. Also, routinely in Govt. douments and land records voiced letters g and j are turned into Tamil consonants. For example, கெசம் is widely used both in Govt. and in public and in Universities. The question is how to give the Character name for கெசம், It can be KECAM to be consisitent with the statement of D. Anderson et al., reco,
"While we agree that it is not clear what Tamil transcription standard is being referred to, the Tamil Virtual Academy's recommendation seems to be based on dictionary forms, such as those in the University of Madras Tamil Lexicon, where initial c is the preferred spelling (instead of initial s), initial p (instead of initial b), medial k (instead of medial g), medial t (instead of medial d), and the long vowels spelled out. By relying on the entries in a standardized reference (such as a dictionary), the spelling in the Unicode Standard will be as consistent as possible. "

Tamil Nadu experts' feedback: கெசம் is used for this character in Govt. and public, and either KECAM or KESAM can be employed in Unicode Character Names. To be consisitent thorough out for Tamil BMP and SMP blocks, they recommend using KECAM.

We can send an Updated document with revised Names and Annotations to the Editorial Committee. The governing principle will be, as stated by Deborah Anderson et all .,
"While we agree that it is not clear what Tamil transcription standard is being referred to, the Tamil Virtual Academy's recommendation seems to be based on dictionary forms, such as those in the University of Madras Tamil Lexicon, where initial c is the preferred spelling (instead of initial s), initial p (instead of initial b), medial k (instead of medial g), medial t (instead of medial d), and the long vowels spelled out. By relying on the entries in a standardized reference (such as a dictionary), the spelling in the Unicode Standard will be as consistent as possible. "

Thanks,

Naga Ganesan