

Neo-Brahmi Generation Panel:

## Analysis of comments for Tamil script LGR Proposal for the Root Zone

Revision: June 30, 2019

Neo-Brahmi Generation Panel (NBGP) published the Tamil script LGR Proposal for the Root Zone for [public comment](#) on 25 September 2018. This document is an additional document of the public comment [report](#), collecting NBGP analyses as well as the concluded responses. There is 1 (one) comment submission. The analysis is as follow:

No.	1	From	Liang Hai
Subject	A quick review of the Tamil proposal		
Comment	§3, “tholkəppiyəm” and many other similar cases throughout the document: Use a common transliteration scheme (say, ISO 15919) consistently.		
NBGP Analysis	Agree.		
NBGP Response	Update the proposal as per the comment.		
Comment	§3.1, “The image below shows how vaṭṭeluttu got transformed as Tamil letters.”: The introduction below Figure 1 says the image is about Tamil Brahmi diverging to Vaṭṭeluttu and Tamil.		
NBGP Analysis	Agree.		
NBGP Response	Update the proposal as per the comment.		
Comment	§3.3.1, “It should also be noted that as per Tamil traditional grammar ... in Tamil Traditional grammar.”: This is not a contrast between the Tamil traditional grammar and Unicode’s terminologies, but simply not clearly distinguishing a “consonantal sound/phoneme” and a “consonantal letter/grapheme” (Indic consonantal graphemes don’t necessarily represent pure consonantal phonemes) in common discussions. Eg, the sentence “The Unicode Consonant set of Tamil comprises the following characters” can be		

	rephrased to “Tamil consonantal letters are encoded as the following Unicode characters”, if the authors want to make the distinction clear.
NBGP Analysis	This point has been made to clarify that NBGP has not blindly considered the nomenclature of Tamil consonants and it also has a good understanding about the vowel consonant which is a traditional Tamil grammar term. However NBGP also agrees to change the phrase as suggested by the reviewer.
NBGP Response	Update the proposal as per the comment.
Comment	§3.3.1, Table 3: Oh c’mon, can we at least clean up the format so phonetic transcriptions are legible?
NBGP Analysis	Agree.
NBGP Response	Update the proposal as per the comment.
Comment	§3.3.1, Table 3: Oh c’mon, can we at least clean up the format so phonetic transcriptions are legible?
NBGP Analysis	Agree.
NBGP Response	Update the proposal as per the comment.
Comment	§3.3.2, Example 1 and 2: Use a clear format if character names are used, eg, <TAMIL LETTER KA, TAMIL SIGN VIRAMA, TAMIL LETTER SSA>.
NBGP Analysis	This is based on devanagari document format. NBGP considered devanagari format as a template to maintain the uniformity across the others LGR documents. Hence The NBGP does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§3.3.4, “The Visarga is also used in Tamil and represents a sound very close to /k̐/”: “k̐” is the ISO 15919 transliteration for ː, not a phonetic transcription. This sentence effectively means “ː represents a sound very close to /ː/”.

NBGP Analysis	The NBGP does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§4.1.2.4, 'AU LENGTH MARK “ௌ” (U+0BD7) is a character in Tamil which has been added to Unicode and is very rarely used in Modern Tamil.' This character is encoded in Unicode for technical reasons, and is part of ஓௌ and ெௌ's canonical decompositions (basically, because U+0BCC TAMIL VOWEL SIGN AU has its canonical decomposition as <U+0BC6 TAMIL VOWEL SIGN E, U+0BD7 TAMIL AU LENGTH MARK>, U+0BD7 is technically as much used as U+0BCC is), therefore the rationale of “very rarely used in Modern Tamil” is inappropriate although the recommendation of excluding it on the exclusion-principle-level is okay (the variation-level is probably better?). Instead probably state the character is not used in valid Tamil text that is in Unicode Normalization Form C (NFC).
NBGP Analysis	NBGP agrees with the suggestion and thanking the reviewer for his inputs on AU LENGTH MARK.
NBGP Response	Update the proposal as per the comment.
Comment	§5.2, “for Tamil Language that NBGP has considered as given in 3.2”: Isn't it stated in §3.2 that “they have not been considered in the present analysis”?
NBGP Analysis	This is just mentioned to sync the connectivity of the document with sections 3.2 and 5.2. The NBGP does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§5.2, Table 5: Should note the “Indic syllabic category” column is not about the Unicode character property of the same name.
NBGP Analysis	Agree.
NBGP Response	Update the text from “Indic syllabic category” to “Category”.

Comment	§5.2.1, Table 6a: Are those “=” in the second column intended?
NBGP Analysis	It is intended as per the uniformity of the document.
NBGP Response	No action required.
Comment	§5.5, “... in the form of variables”: These are not variables but notation.
NBGP Analysis	A variable is something which acts as a placeholder for multiple entities with the same properties as intended by the creator of the variable. The entities mentioned are also being viewed as such. The Tamil Team does not see the need to change as suggested.
NBGP Response	No action required.
Comment	<p>§5.5.4, “3. A sequence of consonants ...”: As it’s already stated in 5.5 that this section is about “Akshar formation”, it’s unclear why the authors are bringing phonetic syllables into the discussion. The so called “CHCHC” is just 3 separate akshars, &lt;CH, CH, C&gt;, although phonetically they belong to the same syllable. As Tamil doesn’t ever have pre-base structures written before (to the left of) a pulli-ed consonant letter, it doesn’t need to follow Devanagari’s practice and can safely simplify its akshar formation logic to allow only a single base consonant — the special cases śrī and kṣV should be discussed as special cases (because they’re really special, unlike anything else in Tamil).</p> <p>- The akshar formation pattern is thus simplified to `(C[M] V)[X]   C[H]` (note V is just a special case of CM when C is zero), or expanded to `C[M][X]   V[X]   C[H]` for the sake of being more comprehensible. But note this analysis hasn’t taken the akshar-preceding visarga/aytham into consideration.</p>
NBGP Analysis	The NBGP does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§6.1.3: My first impression is this should be “blocked” because <sa, virama, ra, vowel sign i> and <sha, virama, ra, vowel sign i> are indistinguishable, but I’m aware that I’m not familiar enough with the concept of “allocatable” vs “blocked”. If the intention is to explicitly allow the applicant to make both encodings aliases to each other (while still

	blocking other applicants from applying for the variant) so users can access the same domain no matter which encoding they use, then it's good.
NBGP Analysis	It was recommended by IP during the review rounds that they should be treated as allocatable, and this decision was taken after the decent amount of discussions within NBGP and it does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§6.3: Should note such cases are already eliminated by the NFC requirement of IDNA2008?
NBGP Analysis	Both cases are IDNA2008. The NBGP does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§6.4, Table 21: The bottom-right cell has a wrong rendering of the Malayalam text.
NBGP Analysis	Agree.
NBGP Response	Update the proposal as per the comment.
Comment	§6.5.2 Allocatable variants: See the comment above for 6.1.3. (I'm not confident about my understanding of "allocatable" and "blocked" ...)
NBGP Analysis	As discussed in 6.1.3 The NBGP does not see the need to change as suggested.
NBGP Response	No action required.
Comment	§7, "... by all the languages mentioned in section 3.2 ...": See the comment above for 5.2, "... for Tamil Language that NBGP has considered as given in 3.2".
NBGP Analysis	Agree.

NBGP Response	Update the proposal as per the comment.
Comment	§7, “Below are the specific WLE rules”: See comment above for §3.3.4, “To facilitate this modern usage apart from barring Visarga – Visarga combination ..”. Also, it’s unclear how the authors achieved this set of rules from §5.5’s analysis. Considering the consonant-modifier visarga, the akshar pattern should be `[X]C[M][X]   V[X]   [X]C[H]` (or even allowing X to precede V, if there’s attestation).
NBGP Analysis	The Authors of this document are well versed with the ISCII standard and the C-DAC GIST IDN Policy documents from where this comprehensible pattern is taken and suggested. The Section 7 is meant to be simplified version of the same with additional bounds that the LGR procedure puts. The rules given in Section 7 have been specifically made simple to be “comprehensible” even to a non-technical user. It is unfortunate that the commenter could not “comprehend” the same.
NBGP Response	No action required.
Comment	§11 Appendix A, Table 22: Based on the same level of similarity, the following pairs (and probably more) should also be considered: U+0B89 TAMIL LETTER U and U+0D09 MALAYALAM LETTER U, U+0BB5 TAMIL LETTER VA and U+0D35 MALAYALAM LETTER VA, U+0BB7 TAMIL LETTER SSA and U+0D37 MALAYALAM LETTER SSA.
NBGP Analysis	NBGP has analyzed all characters between Tamil and Malayalam for the variant code points / similar code points and distinguishable code points and agreed that there is only one similar set. The NBGP does not see the need to change as suggested.
NBGP Response	No action required.