

Response to L2/19-345: *Alternative encodings for Malayalam “nta”*

Cibu C Johny
 cibucj@gmail.com
 2019 Oct 06

I am curious to know the reference for the phonetic analysis described in section *A chillu-less analysis* in the proposal L2/19-345. How can a phonetic analysis be the basis for an important double encoding decision?

Actually, Kevin Suryan and I might have come up with this sequence in 2004 Sept, as we created the initial set of font and input tool for the Malayalam implementation of Unicode. Unicode standard at that stage did not explicitly define the sequence for റ്റ (⟨⟨chillu N, subscript RRA⟩⟩ /ntʌ/). Anycase, the sequence implied by this particular analysis is an artifact of the evolution of Unicode for Malayalam; it is not grounded in any prior writing traditions or academic literature.

CHILLU N is phonetically realized only as alveolar /n/, while NA can be dental /ɳ/ or alveolar /n/. In Malayalam, dental /ɳ/ and alveolar /n/ are not allophones as implied in the proposal. So using <NA, VIRAMA> for CHILLU N is not phonetically accurate. Moreover, if you show the visual റ്റ (⟨⟨NA, visual VIRAMA, RRA⟩⟩) to a native user (who is unaware of Unicode particulars), they will not identify it as റ്റ (⟨⟨chillu N, subscript RRA⟩⟩ /ntʌ/); instead, they would read it as /nəɳ/.

This proposal does not address the remaining chillu conjuncts described in [L2/19-086R](#). It also does not address the legacy sequence supported by MS Windows <NA, VIRAMA, ZWJ, RRA> for റ്റ (⟨⟨chillu N, subscript RRA⟩⟩).

I am not sure how this proposal is going to solve the issue of inadequate support for <CHILLU N, VIRAMA, RRA>, without explicitly rescinding this sequence. Double encoding for റ്റ (⟨⟨chillu N, subscript RRA⟩⟩) is not going to solve any issue, if not, making the issue more acute. Double encoding is never a desirable quality for Unicode. So the decision should not be taken lightly or hastily. It needs to be clearly thought through, probably through a PRI.

That said, I sympathize with the author’s intention, a font developer should be made aware of the prevailing usage of <NA, VIRAMA, RRA> and <NA, VIRAMA, ZWJ, RRA> sequences for റ്റ (⟨⟨chillu N, subscript RRA⟩⟩). So, there can be some text in the standard, similar to that used for the legacy encoding of the chillus. For example, in the following text can be added to the section *Legacy Chillu Sequences* in page 512 of Unicode version 12.1:

Prior to Unicode 5.2, the encoding of the cluster ᱠᱚ (<<chillu N, subscript RRA>> /nt^/) was not clearly defined. So implementations should be prepared to handle the non-standard legacy usage of <NA, VIRAMA, RRA> and <NA, VIRAMA, ZWJ, RRA> sequences representing this cluster. There is no legacy usage prevalent for the remaining conjuncts involving chillus.

The sequence <NA, VIRAMA, RRA> for ᱠᱚ (<<chillu N, subscript RRA>>) should not be legitimized as an *alternate encoding*; but should be recognized as a prevailing non-standard legacy encoding.