

On the usage of nuktas in Tamil

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2016-Jan-28

Based on Martin Hosken's proposal L2/15-256, the UTC in its Nov 2015 meeting has decided to encode 0BBC TAMIL SIGN NUKTA in the standard to support the single-dot nukta used in the Tamil script for certain minority orthographies. The minutes of that meeting record that based on the same document, 1133C GRANTHA SIGN NUKTA will get Tamil added to its Script Extensions property to support the double-dot nukta used in the Tamil script for other minority orthographies. Ref: <http://www.unicode.org/cgi-bin/GetL2Ref.pl?145-C10>

It is unclear as to why two different approaches are applied to two identical requirements. If it is appropriate to just add Tamil to the Script Extensions of 1133C GRANTHA SIGN NUKTA to support the double-dot nukta usage, it seems equally appropriate to just do the same to 093C DEVANAGARI SIGN NUKTA to support the single-dot nukta usage.

On the other hand, if there is a problem with using the Devanagari nukta with Tamil text then the same would apply to the Grantha nukta as well. However, it would seem that we are well beyond the point where characters need to all be in the same block to behave well with one another in text shaping processes. So the earlier suggestion of borrowing both nuktas from other blocks seems the more practical one for a variety of reasons.

This need not be construed to be against encoding any future script-specific nukta characters if such native and long-standing usage is evidenced in attestations (as recently we have seen some Gujarati nukta characters encoded). The present approach is recommended only for Tamil since apparently this usage of nuktas is a recent (even computer-age?) innovation and just "borrowing" existing characters is much simpler than encoding new ones.

That Devanagari is not directly related to Tamil historically (but Grantha is) is irrelevant from a technical point of view. Already Devanagari combining characters udatta and anudatta 0951 and 0952 have had their script changed to "Inherited" for widespread use with all Indic scripts that have been used for Vedic (including Tamil if needed). Likewise the "Devanagari" dandas 0964 and 0965 are used in multiple Indic scripts including Tamil.

As such, it is recommended to rescind the approved character 0BBC TAMIL SIGN NUKTA and instead just add Tamil to the Script Extensions property of 093C DEVANAGARI SIGN NUKTA such as added to 1133C GRANTHA SIGN NUKTA. Appropriate annotations pointing to these characters should be added either in the Tamil code chart or the Tamil description chapter of TUS or both.