L2/10-405

On the GC of 1134D GRANTHA SIGN VIRAMA

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

2010-Sep-17

In the Grantha script, which is in use for representing the Sanskrit language, three different ways of representing vowelless consonants are seen to exist – three different "virama forms" so to speak:

ட் ட் ட் த் த் த ந் ந் ந

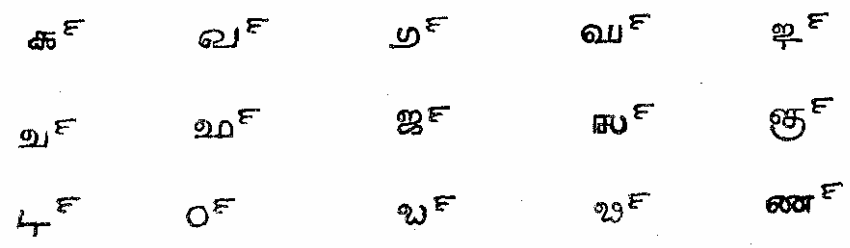
In my Grantha proposal L2/09-372 §5.3 and §6.2 I have discussed these forms and referred to them as the spacing, touching and ligated virama forms (in the order shown above).

Now since these forms are all semantically identical, no difference in encoded representation is warranted. The suggestion in L2/09-372 (especially §6.2.1) and in my other previous documents of the need of a distinct LIGATING VIRAMA was based on a purely hypothetical need to distinguish between the three forms in plaintext. However, such an encoding of a distinct character is fraught with problems and therefore it is best that only a single VIRAMA, with the help of appropriate glyph substitution tables, is used to represent all three forms as per the desired orthographic style, as recommended in L2/10-167 p 4.

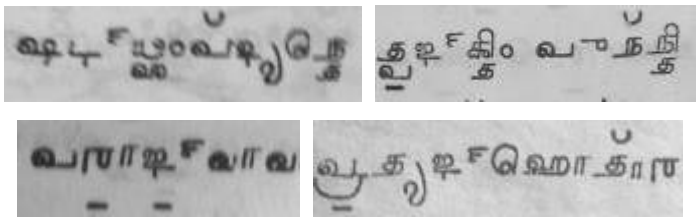
Thus 1134D GRANTHA SIGN VIRAMA is the only virama character to be encoded for Grantha. Now it is to be determined what the GC of this character should be.

When it either touches or ligates with its base consonant it does not cause any advance width – that is clear. However, the question is about the default virama – that which does not come into contact with its base at all. (The touching and ligating virama forms may be considered glyphically joined/fused variants on the spacing forms, but obviously the converse is not possible. Thus the spacing form is the "default".)

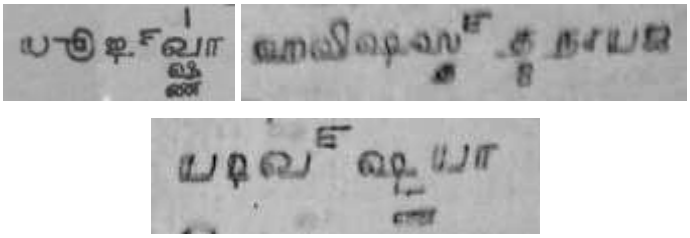
That this character is spacing is quite visible plainly from the most basic Grantha books like primers which are intended for teaching the script. I have provided sufficient samples for this with proper references on p 21 of my proposal as reproduced here:



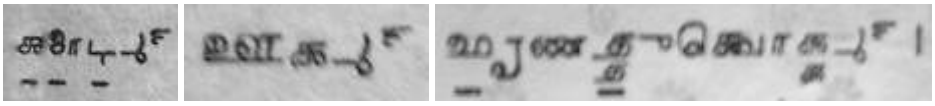
The above sample shows the Grantha virama in isolation with consonants. The fact that it takes a positive advance width and hence causes the cursor to advance in the writing direction (LTR, obviously) is much clearer when the explicit virama occurs in the middle of consonant clusters (as a fallback for conjoining forms or by specific desire of the writer) as shown in p 23 of my proposal: (for references see the pages mentioned)



Further in p 27 and p 29 we have:



A further strong (may I say unshakeable) argument in favour of the virama being spacing is the fact that the repha will intervene between the virama and its consonant, as on p 30:



Thus I think that the virama should considered spacing and hence given GC=Mc.

The argument in favour of GC=Mn is only that no other Indic scripts have GC=Mc for their virama characters at all. All currently encoded Indic virama characters have GC=Mn. In fact, as of Unicode 6.0 the only characters with CCC=9 and GC=Mc are from Balinese, Sundanese, Batak, Rejang and Javanese – all South East Asian, no Indic scripts. However, this is a weak argument, in my opinion, as each script may justifiably have distinct characteristics that are not present in sister scripts.

I note that 0D4D MALAYALAM SIGN VIRAMA, which is perhaps positionally the closest relative of 1134D GRANTHA SIGN VIRAMA, being placed like it to the top-right of a consonant, has GC=Mn. Malayalam fonts do show a very small advance width for it, though. In Grantha, however, the advance width is not "small" by any means, as seen in the samples shown above, especially those involving consonant clusters and the repha.

Thus I feel 1134D GRANTHA SIGN VIRAMA should get GC=Mc and not GC=Mn. The UTC must decide on the Indic_Matra_Category property for this character, though.

-o-o-o-