

Finalizing the Grantha virama model

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

2013-Nov-10

This document is basically a summary of my older document L2/10-404 with the same title, especially of its §3.4 on p 23, and recommends practically the same virama model for Grantha as that document did. However, this document also draws upon the precedent of parallel cases in other Indic scripts that have come to light in the intervening years.

§1. Parallel cases elsewhere in Indic

§1.1. Telugu

In the Telugu script, the consonant NA న has two vowelless forms: ఞ and న్ . My document L2/11-409 presented evidence for this, pointed out that it is merely a stylistic choice as to which of the two forms is used as there is no semantic difference, and recommended the following encoding model. This model was accepted at the 2012 Feb UTC meeting.

Fonts are free to render the isolate sequence NA + VIRAMA as desired:

Old Style:	NA + VIRAMA	ఞ
New Style:	NA + VIRAMA	న్

When the sequence is not isolate but is followed by another consonant, NA + VIRAMA + C2, the rendering would of course involve (ligatures or) conjoining forms:

All styles:	NA + VIRAMA + DA	ఞ్
-------------	------------------	----

However, if the combining behaviour is broken up by inserting a ZWNJ before the second consonant, then again the NA + VIRAMA would be as good as isolate and would again be rendered as per the style:

Old Style:	NA + VIRAMA + ZWNJ + DA	ఞ్‌
New Style:	NA + VIRAMA + ZWNJ + DA	న్‌

While the same text passage in handwriting or printing may occasionally show both forms, there is no need to represent the distinction in plaintext. In rich-text, smart font features may be used to specify a particular form in desired positions if needed.

§1.2. Kannada

My recent document L2/13-228 provides evidence for the same orthographic feature in the closely related script Kannada as well and recommends the same model there:

Isolate:	NA + VIRAMA	ಃ/ನ್
Isolated by ZWNJ:	NA + VIRAMA + ZWNJ + DA	ಃದ/ನ್ದ
Non-isolate:	NA + VIRAMA + DA	ನ್ದ

§1.3. Bhaikshuki

Anshuman Pandey’s Bhaikshuki proposal L2/13-194 (p 4) identifies special forms of vowelless TA, NA and MA in Bhaikshuki as well (the ones on the right below):

< ॐ TA, ॐ VIRAMA >	ॐ	ॐ
< ॐ NA, ॐ VIRAMA >	ॐ	ॐ
< ॐ MA, ॐ VIRAMA >	ॐ	ॐ

Again, there is no semantic distinction and no need to distinguish in plain encoded text. A font is to provide an acceptable default representation of the sequence C + VIRAMA as per a chosen typographic style. Smart font features are again sufficient to control appearance of individual cases in rich-text.

Thus so long as there is semantic equivalence of multiple vowelless forms of a consonant and no other consideration prevents their being unified in plain text, such a model is quite sufficient to cater to such multiple vowelless forms* in Indic.

§2. Similar model for Grantha

In Grantha, many consonants have upto three different vowelless forms – where the virama stands apart from the consonant as in ॐ and ॐ (for KA and TTA), where it superficially connects to the consonant as in ॐ and ॐ, and where it is considered to totally fuse with the consonant as in ॐ and ॐ. Just as in Telugu, Kannada and Bhaikshuki, it is merely a stylistic consideration as to which form is used and they are all mutually entirely equivalent.

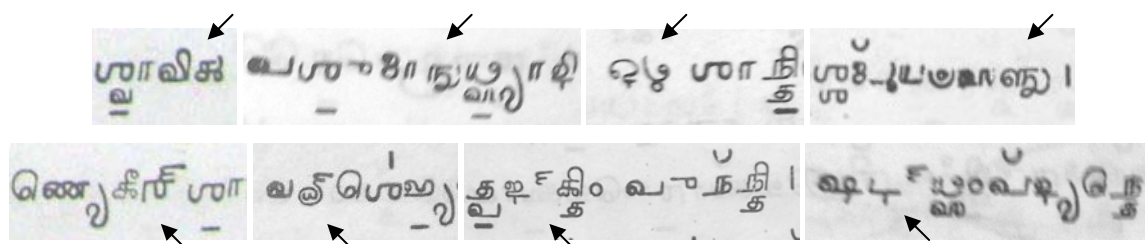
The mutual equivalence of these forms has been demonstrated in my Grantha proposal L2/09-372 §5.3.2 (p 22). The manuscripts experts who attended the GOI meeting

* Note that in Malayalam, there is no semantic equivalence between the chillu and chandrakala forms, since the latter may also denote the “short u” of Malayalam known as samvruthokaram. In Bengali, the khanda-TA is considered as a distinct grapheme of the script (substituting TA-virama instead is considered a “spelling mistake”). Thus the above rule does not apply there and hence the special vowelless forms in those scripts cannot be treated as equivalent to CONSONANT + VIRAMA in plain-text.

on Grantha conducted 2010-Sep-06 also agreed that these should all be treated equivalently as recorded in L2/10-409 pp 3-4.

Therefore the same model applicable to Telugu, Kannada and Bhaikshuki is appropriate for Grantha also: the isolate sequence C + VIRAMA should be freely rendered as the desired vowelless form of a particular consonant in a particular orthographic style.

For example, in the prevalent orthographic style seen in modern printings, TA, NA and MA and N·NA consistently fuse with the virama, RA and LA superficially connect with it, and for all other consonants the virama stands apart, as L2/09-372 p 35 shows:



Therefore it is sufficient that when either the fused or superficially connected form is required, a font substitutes the sequence C + VIRAMA by the appropriate glyph, and in all other cases, it renders the sequence without substitution:

Fused:	TA + VIRAMA	த + ்	→ ச
	NA + VIRAMA	ந + ்	→ று
	MA + VIRAMA	ம + ்	→ ழ
	N·NA + VIRAMA	ஊ + ்	→ ளு
Connected:	RA + VIRAMA	ர + ்	→ ர்
	LA + VIRAMA	ல + ்	→ ல்
Unconnected:	KA + VIRAMA	க + ்	→ க்
	TTA + VIRAMA	ட + ்	→ ட்

When the sequence C + VIRAMA is not isolate but followed by another consonant as in C + VIRAMA + C2, the rendering would be as either a ligature of C and C2 or involve conjoining forms (i.e. if C = RA then the reph is formed, else C2 takes a sub-base form):

TA + VIRAMA + SA	த + ் + வு	→	த்
NA + VIRAMA + DA	ந + ் + டி	→	ந்
RA + VIRAMA + HA	ர + ் + ஹ	→	ரஹ்
LA + VIRAMA + SHA	ல + ் + ஸ	→	லஸ்த்
TTA + VIRAMA + VA	ட் + ் + வ	→	ட்வ

Again, if the combining behaviour is broken up by inserting a ZWNJ before the second consonant, then again the C + VIRAMA sequence would be as good as isolate and would be rendered as per the desired orthographic style:

Fused:	TA + VIRAMA + ZWNJ + SA	த + ் + ZWNJ + வு	→	த்வு
	NA + VIRAMA + ZWNJ + DA	ந + ் + ZWNJ + டி	→	ந்தி
Connected:	RA + VIRAMA + ZWNJ + HA	ர + ் + ZWNJ + ஹ	→	ரஹ்
	LA + VIRAMA + ZWNJ + SHA	ல + ் + ZWNJ + ஸ	→	லஸ்த்
Unconnected:	TTA + VIRAMA + ZWNJ + VA	ட் + ் + ZWNJ + வ	→	ட்வ

§3. Summary

In summary, it is recommended that Grantha should follow the same virama model as found useful for Telugu, Kannada and Bhaikshuki:

- 1) The basic sequence C + VIRAMA when isolate (either naturally or due to a following ZWNJ) will be displayed by the font as the default vowelless form of C in the desired typographic style:

C + VIRAMA (isolate) → DEFAULT VOWELLESS FORM OF C

C1 + VIRAMA + ZWNJ + C2 → DEFAULT VOWELLESS FORM OF C1 + C2

- 2) When CONSONANT + VIRAMA occurs as part of a consonant cluster, the normal Indic behaviour of consonant ligatures or conjoining forms will occur:

C1 + VIRAMA + C2 → Ligatures or Conjoining Forms

- 3) Smart font features may be used to request a particular vowelless form in rich text.

It is requested that the Unicode chapter on Grantha include these guidelines.

-o-o-o-