Handling variation in vowelless consonant forms in Grantha

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

2014-Oct-29

§1. Introduction

This is a response to Ganesan's L2/14-162 and L2/14-279 and the review committee's L2/14-268 §1. Please also refer to my earlier 2-page document L2/14-110 responding to Ganesan's earlier document L2/14-097.

Ganesan previously claimed semantic distinction for the Grantha vowelless forms which he calls "chillus". This has been entirely refuted by multiple documents from myself, the native user community and finally by the high-level scholarly committee appointed by the GoI as seen in L2/10-409 pp 3, 4. (For further details see my earlier doc.) Hence where he first wished to encode a chillu marker, now he advocates the use of ZWJ.

§1.1. No widespread demand

Now I wish to impress upon the UTC the fact that there is no demand from the broad Grantha user community for mechanisms to represent *in plaintext* the variation in the Grantha vowelless forms. I do not understand the reasons for Ganesan's lone persistence in this issue but in any case it should not be interpreted as representing a wider demand.

I am personally aware of the requirements of the user community because I am one of the few thousands of native Grantha users here in Tamil Nadu who are chiefly Vedic/Sanskrit scholars and practitioners like myself and I am in continuous contact with these peers of mine. I am also in contact with computer-savvy research scholars handling Grantha manuscripts and I have not heard any complaints from them re the present issue (or any other) either. Finally, the scholars who formed part of the GoI-appointed scholarly committee are eminently representative of both the native and research user communities and they have in fact emphasized that the variation in the vowelless forms should not be handled at the encoding level but only at the font level (see L2/10-409 pp 1, 3 and 4).

§1.2. Potential mechanism

Notwithstanding all this, even if it is perceived as useful *in principle* to have a mechanism for representing these in plaintext, it is important that any such mechanism should:

1) conform to the semantic equivalence of the Grantha vowelless forms
2) not conflict with the existing Unicode model for other Indic scripts

I shall demonstrate that the suggestion to use C + VIRAMA + ZWJ to denote the vowelless forms under question is not in line with the above principles, especially the second.

§2. Problems with the suggestion of C + VIRAMA + ZWJ

Ganesan has been suggesting the usage of C + VIRAMA + ZWJ for what he calls Grantha chillus i.e. the "fused" virama forms க் *k*, ன் *ṅ*, ட் *ṭ* etc. In L2/14-279, he quotes the pan-Indic virama model based on Peter Constable's PRI 37 doc L2/04-279 (a numerical coincidence!) and claims to follow the same model in his proposal[1].

§2.1. Fused virama forms are not conjoining forms

Now as per the accepted model as outlined in Peter's document, in Indic C + VIRAMA + ZWJ is used to select the C1-conjoining form of the consonant C. A conjoining form is an alternate form that a consonant takes when it is part of a consonant cluster.

In Grantha, the fused virama forms are freely used whether inside or outside consonant clusters and freely alternate with other virama forms (see L2/09-372 pp 21-22 for many samples) and hence cannot be considered conjoining forms. They are totally independent vowelless forms.

Further, in Grantha, the only attested C1-conjoining form is the reph ர and all other consonants only have C2-conjoining forms as seen in other South Indian scripts like Telugu and Kannada[2]. In these scripts, RA + VIRAMA + ZWJ is the sequence used to request the reph in isolation and in preference over other conjoining behaviours as seen in TUS 7.0 p 481 (p 519 of PDF). The same should be maintained for Grantha as well. Note that it is possible that further research into old manuscripts may turn up archaic C1-conjoining forms for other consonants in Grantha as well. As such, the sequence C + VIRAMA + ZWJ should be restricted to requesting the C1-conjoining forms as in other Indic scripts. And, as illustrated copiously in earlier documents, fused virama forms are *not* conjoining forms. Note also that the C1-conjoining form of *r* i.e. the reph ர is different from its fused virama form viz ற்.

---

[1] While claiming this, he shows the glyphic representation of Devanagari KA + VIRAMA + ZWJ + SSA incorrectly as क्‌ष instead of क्ष. In a document which purports to recommend the correct usage of joiners!

[2] Due to the typewriter-era-induced semantic-chillu situation, one is not able to draw the Malayalam parallel.

§2.2. The role of ZWNJ

Ganesan claims in L2/14-279 p 2 §7 that Peter Constable said at the last UTC meeting that my proposal to use ZWNJ is inconsistent with the prevalent Indic model. I do not know what Peter said or didn't say at the last meeting. I do know what is published publicly:

As per the published standard, the role of ZWNJ has always been only to block the combining behaviour of the preceding and following sequences.

Ref: TUS 7.0 p 438 (p 476 of PDF) (underlining mine):

> *Occasionally, this default behavior is not desired when a dead consonant should be excluded from conjunct formation, in which case the virama sign is visibly rendered. To accomplish this goal, the Unicode Standard adopts the convention of placing the character U+200C ZERO-WIDTH NON-JOINER immediately after the encoded dead consonant that is to be excluded from conjunct formation. In this case, the virama sign is always depicted as appropriate for the consonant to which it is attached.*

TUS 7.0 pp 804-805 (pp 842-843 of PDF):

> ***Joiner and Non-joiner in Indic Scripts.*** *In Indic text, the ZWJ and ZWNJ are used to request particular display forms. A ZWJ after a sequence of consonant plus virama requests what is called a "half-form" of that consonant. A ZWNJ after a sequence of consonant plus virama requests that conjunct formation be interrupted, usually resulting in an explicit virama on that consonant. ...*
>
> ***Implementation Notes.*** *For modern font technologies, ... font vendors should add ZWJ to their ligature mapping tables as appropriate. ... In contrast, ZWNJ will normally have the desired effect naturally for most fonts without any change, as it simply obstructs the normal ligature/cursive connection behavior. As with all other alternate format characters, fonts should use an invisible zero-width glyph for representation of both ZWJ and ZWNJ.*

Thus the ZWNJ is not really part of the *selection* mechanism of a particular presentation form as it is not intended to be used in glyph substitution tables; rather it is used merely to *prevent* any sort of combining behaviour. In a sequence such as C1 + VIRAMA + ZWNJ + C2, the text before and after the ZWNJ will simply be rendered as if the other did not exist.

Thus it is evident that where the Indic virama model recommends the usage of ZWNJ for the so-called level 3 presentation of consonant clusters, it is not partial to any one virama form against another, but only ensures that a conjoining form is *not* displayed.

The Telugu and Kannada sections of TUS 7.0 pp 481 and 485 (pp 519 and 522 of PDF) recommend that the special vowelless forms of NA in those scripts (called *nakāra pollu* in Telugu) are produced within clusters by the usage of ZWNJ. This is also simply based on blocking of the interaction with the following consonant allowing the font to render the CONSONANT + VIRAMA sequence as desired:

> *The character U+200C ZERO-WIDTH NON-JOINER can be used to prevent interaction of this sequence with the following consonants...*

It is the very same blocking behaviour of ZWNJ that I have suggested to employ for Grantha in my brief document L2/14-002 which summarizes my earlier L2/10-404:

C1 + VIRAMA + ZWNJ + C2 → default virama form of C1 + C2

This is again merely a technical restatement of the recommendations of the GoI-appointed scholarly committee in L2/10-409 p 3 that:

> *... these vowelless forms ... should be handled at the font level, not at the encoding level... the rendering ... can be worked out in font software.*

Ganesan objects to handling the variation at the rich-text level – it is not clear why, especially when nowadays it is not even necessary to change fonts but the same font with different "features" enabled can cater to these stylistic variations. In any case, Ganesan's suggestion to use the sequence C + VIRAMA + ZWJ to denote the fused virama forms is against existing Indic practice and hence should not be adopted for Grantha.

§3. Alternate suggestion using Variation Selectors

As a last resort, if one is to absolutely demand a solution, I would recommend this:

C + VIRAMA + VS1 → "Spacing" virama form of C such as క్‌

C + VIRAMA + VS2 → "Touching" virama form of C such as క్

C + VIRAMA + VS3 → "Fused" virama form of C such as క్‍ (if attested)

These sequences are nowhere used in Indic and hence do not conflict. No other Indic script has three different virama forms. Font makers who wish to cater to such a requirement can simply add the relevant mappings to their fonts in the first substitution pass.

-o-o-o-