# Further clarification on Grantha virama-ligatures

Shriramana Sharma – jamadagni-at-gmail-dot-com

2009-Oct-06

## *Brief recap of L2/09-316*

I previously submitted the document L2/09-316 on why a 'chillu marker' (as suggested by Mr Ganesan) should not be encoded for Grantha. This document is intended to further clarify some other aspects of the same issue.

The previous document used the terminology 'Grantha chillu-s' to denote the forms like &#x0C95; &#x0C9C; &#x0C9E; &#x0CB4; etc where the virama ligates with the consonant to produce vowelless consonant forms graphically distinct from the 'overt-virama forms' க௭ ந௭ ஙௗ உௗ etc.

That document took pains to demonstrate that the 'overt-virama forms' and 'chillu forms' in Grantha are semantically non-distinct for Sanskrit and hence entirely interchangeable in usage *unlike in Malayalam* where മനുഷ്യൻ can have a different meaning from മനുഷ്യന്. There is no such semantically contrastive usage in Grantha of 'chillu forms' as against 'overt-virama forms'. Minimal pairs like ஐஐஙௌந� and ஐஐஙௌந௭ have the same meaning. Claims that the usage of 'chillu forms' in Grantha causes change of meaning are baseless and contrary to the nature of the Sanskrit language.

Further, the output of C + CHILLU MARKER where C does not have a 'chillu' in Grantha would be undefined. Fallback to overt-virama forms in such cases would result in security problems since C + VIRAMA would lead to the same. Such arguments speak against the encoding of a 'chillu marker' and are discussed in detail in L2/09-316.

## *'Virama-ligatures'*

On reflection, we realize that it is better to entirely avoid the terminology of 'Grantha chillu' because it misleads readers who are conversant with Malayalam chillu-s by causing them to think that these characters are the direct Grantha parallels *in every way* of Malayalam chillu-s, despite proof to the contrary. Thus we shall use the alternative terminology of 'virama-ligatures' as it gives no meaning other than that the virama ligates with the consonant.

An important distinction between Malayalam chillu-s and Grantha virama ligatures is that Malayalam chillu-s are analysed as conjoining forms (PR 37 p 5) because they occur to a large extent as part of consonant clusters, though they may also denote isolated (terminal) consonants. However, in Grantha, virama-ligatures are *not preferentially* used in consonant clusters. A consonant cluster in Grantha is normally written as a ligature, using sub/post-base conjoining forms, or using a combination of the above. For example:

$$\text{KA} + \text{VIRAMA} + \text{SSA} \rightarrow \text{கூஷ}$$

$$\text{KA} + \text{VIRAMA} + \text{VA} \rightarrow \text{க்வ}$$

1

KA + VIRAMA + YA → க்ᵧ

KA + VIRAMA + SSA + VIRAMA + VA → க்ஷ்ᵥ

KA + VIRAMA + SSA + VIRAMA + YA → க்ஷ்ᵧ

KA + VIRAMA + SSA + VIRAMA + VA + VIRAMA + YA → க்ஷ்ᵥᵧ

*Thus neither virama-ligatures nor overt-virama forms are seen as part of normal consonant clusters in Grantha.* They occur in the natural writing of consonant clusters only when there are excessive stacking elements causing the first consonant(s) to be pushed out of the cluster and written separately:

NGA + VIRAMA + GA + VIRAMA + DHA + VIRAMA + VA + VOWEL SIGN E → ங்க்ᵈᵥ

TA + VIRAMA + SA + VIRAMA + TA + VIRAMA + NA + VOWEL SIGN E → த்ஸ்ᵗ்ன

But even here *the role of the overt-virama and virama-ligature forms is only as fallbacks to compensate for the absence of conjoining forms placed too many stacking levels below the base character.* Such fallbacks due to over-stacking not being supported occur in other Indic scripts too, such as when writing the (Vedic) Sanskrit sequence vyabhavathsphyaḥ in Malayalam: വ്യഭവഡ്സ്പ്യഃ whereas in Grantha: வ்யஎவஎஷ்ப்யஃ

Just as one realizes that � is not a combining form but only a fallback, one must realize that even in Grantha, though virama-ligatures (and overt-virama forms) may be seen (though only very rarely) as part of consonant clusters, they are used there only as fallbacks and *hence cannot be considered as conjoining forms* or 'half-forms'.

Of course, just as in all Indic scripts, one may opt to break consonants out of the consonant clusters and display them separately, in which case also Grantha consonants would take virama-ligatures, but it is obvious that such optional breaking off does not constitute default behaviour and is not sufficient justification to consider such forms as conjoining forms, for it would mean that in Devanagari क् is a conjoining form since it is used to present the "broken" form of क्ष as क्‍ष.

Thus Grantha virama-ligatures are not congruents of Malayalam chillu-s and hence any arguments that were put forth for Malayalam chillu-s to be separately encoded do not apply to Grantha virama ligatures.

## *Bengali Khanda TA – another non-parallel*

I have also examined in detail the arguments that were put forth for the separate encoding for the Bengali Khanda TA and satisfied myself that none of those arguments apply to Grantha either. These are summarised as follows:

No native users of Grantha/Sanskrit recognize Grantha virama-ligatures as distinct graphemes. They are merely ligated forms of consonants with virama, fully equivalent to non-ligated forms, just as क्ष is fully equivalent to क्‍ष. No dictionaries of Sanskrit in Grantha

list these as distinct graphemes. (The very thought is absurd to me, since the question would immediately arise as to what happens with Sanskrit dictionaries in other scripts.)

There is absolutely no need to treat Grantha virama-ligatures as distinct from non-ligated forms in text processing, just as there is no *real* need to treat क्ष as distinct from क्ष or कृष for such text processes as searching. In fact, there is a need to treat virama ligatures as identical to non-ligated forms for text processing since the same Sanskrit semantic content may be presented in writing either with or without a virama ligature.

Therefore, none of the reasons Mr Ganesan provides, or those that were provided for atomically encoding Malayalam chillu-s or Bengali Khanda TA can apply to Grantha.

## *The existence of three different virama forms in Grantha*

In fact, I have come to the conclusion that there are other reasons to encode a separate *ligating virama* (NOT a "chillu marker" since the term is both unknown to native Grantha users and also misleading, as shown above) but which is canonically decomposable to the regular virama. Based on these reasons I have recently added a "ligating virama" to my proposal, BUT importantly provided a decomposition to the regular virama.

In §5.3.2 of my proposal I have shown further sufficient proof that the virama ligatures of Grantha are semantically identical to the unligated forms over and above what I provided in L2/09-316. The endorsement from Dr Shastri in §17 also confirms this. Thus there should be no doubt whatsoever regarding the semantic equivalence of Grantha virama ligatures to unligated forms. This is why I have provided a decomposition.

As to the reasons of encoding a ligating virama distinct from the regular virama, it is chiefly the existence of three different virama forms in Grantha, as shown in §5.3.1 of my proposal. I had previously not considered the behavioural difference between the spacing and the touching forms, and only when my former co-author Elmar Kniprath argued that the virama should be non-spacing keeping in mind these forms (which he has used as default in his e-Grantamil font) did I realize that the default virama is spacing but that other forms of the virama are non-spacing, and this leads to behavioural variation as shown in §5.4.6 of my proposal.

The repha positions itself between the consonant and its spacing virama, whereas it cannot do so between the consonant and the touching or ligating virama.

## *The need for a ligating virama*

If it can be validly argued that there is a need to represent in Devanagari plaintext the distinction between क्ष, क्ष and कृष then an analogous argument can be certainly made for Grantha to distinguish in plaintext ⎣ˉ, ⎣ and ⎣.

I then realized that using joiners as for Bengali vowel signs cannot satisfactorily represent this three fold distinction. The reasons for this are as follows:

In the first place, an important constraint I place on any model for supporting the various Grantha virama forms is that the default sequence C + VIRAMA should be able to

represent any of the virama forms as per orthographic stylistic choice of the user. Just as a Bengali user should not have to consciously input a character called a "ligating vowel sign" different from the "non-ligating vowel sign" just to achieve the default representation of the syllable 'gu' in old-style orthography, and then input a non-ligating vowel sign for those consonants that do not ligate, similarly Grantha users should not have to input a different sequence (using an "ordinary virama") for the spacing virama form (which is the default as it is most often seen in contemporary Grantha printings compared to the other forms) and then a different sequence (involving a "chillu marker") for the ligated virama forms to achieve the default representation of vowelless consonants in his chosen style.

In fact, as I have shown in §6.2.1 of my proposal, the chosen orthographic style of a user or publisher can involve choosing from all three virama forms. Then users would have to input three different sequences as per the consonant involved to get the default vowelless representation for each consonant in a single writing style.

This certainly is not an ideal situation and I refuse to burden Grantha users with it, especially seeing as users of all other Indic scripts never have to input anything other than the plain vanilla sequences involving consonants, vowels, vowel signs and virama to get the default appearance of their script.

Therefore the sequence C + VIRAMA should be allowed to represent any virama form as per the font maker's == user's choice. Therefore no fixed definition should be placed for this sequence.

With this restriction, I find that the plain use of ZWJ/ZWNJ can only represent a two-way distinction as in Bengali vowel signs with ZWJ being used unambiguously for ligated forms and ZWNJ being used for unligated ones.

A suggestion would be to use a three-way distinction between ZWJ, ZWNJ and a combination of these two. In particular, the following suggestion is possible:

| | | |
|---|---|---|
| C + ZWJ + VIRAMA | ---> | Ligated virama form |
| C + ZWNJ + VIRAMA | ---> | Spacing virama form |
| C + ZWJ + ZWNJ + ZWJ + VIRAMA | ---> | Touching virama form |

This model is obviously adapted from the model prescribed for Arabic in TUS chap 16.2 (except for that it was impossible to define the use of ZWJ for ligatures due to backward compatibility considerations). While the first and second sequences are logically allright, the third sequence does not work out properly. This is because, unlike in Arabic, the two characters involved, Grantha consonants and Grantha virama, do not have individual cursively joining forms the glyphs for which may be substituted for C + ZWJ and ZWJ + VIRAMA with the intervening ZWNJ merely preventing ligatures. In particular, the consonant does nothing at all in the touching virama forms and even the virama takes a cursively connecting form only for the consonants TTA and LLA as seen in page 24 of my proposal. Therefore, to correctly produce a touching virama form, both the consonant and

virama glyphs must be substituted by a single touching virama form glyph. This is inadmissible as it would involve shaping across a ZWNJ and involve serious changes to the way rendering engines work. This is analogous to felling a banyan tree just to pluck a leaf.

I cannot think of any other models using no new ligating virama and using ZWJ/ZWNJ only which can satisfactorily represent the three-way distinction of Grantha virama forms.

Therefore I come to the conclusion that it is necessarily to encode a separate ligating virama which should nevertheless be canonically decomposable to the regular virama in order to maintain semantic equivalence.

Using this extra virama, a satisfactory model can be developed to denote the three-way distinction and yet retain the freedom of the plain sequence C + VIRAMA to cater to the users' default desired representation. This model is as follows:

| | | |
|---|---|---|
| C + VIRAMA | ---> | Default desired virama form |
| C + ZWNJ + VIRAMA | ---> | Spacing virama form |
| C + LIGATING VIRAMA | ---> | Touching virama form |
| C + ZWJ + LIGATING VIRAMA | ---> | Ligating virama form |

On the choices of the sequences, my justifications are as follows:

The default virama is defined to be spacing. Applying it directly to the consonant will produce any of the three forms as per the font. Hence it is necessary to use ZWNJ to prevent any possible connection between the consonant and the default virama to unambiguously represent a spacing virama form.

The extra virama is labeled a "ligating virama" though I make a distinction between ligating and cursively connecting aka touching forms since the term ligation is often used for both cursively connection and "true ligation" involving a "closer bonding" between the two connecting components.

This ligating virama should on simple application produce a touching == cursively connecting virama form. The ZWJ should then be used to promote this connection to the true ligating form. (This is as desired for Arabic but was impossible to implement due to backward compatibility considerations.)

To keep the meaning of the word "ligating" consistent, one may suggest that C + LIGATING VIRAMA should produce a full ligature. If so, then how would one achieve a cursive connection? It is not admissible to use ZWNJ (logical opposite to ZWJ) for this purpose since though TUS chap 16.2 says that "ZWNJ requests the lowest possible connection" and the lowest possible connection between a consonant and the ligating virama is the touching form, this function of the ZWNJ is implemented by merely allowing the ZWNJ to obstruct connection between the preceding and succeeding sequences as TUS p 540 says: "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." Because of this, most vendors do not cause rendering to occur across the ZWNJ, if my understanding is correct.

The only other solution is to encode yet a third virama, which would have a "touching" effect. This third virama should also be decomposable to the first (default) virama. In my opinion, this is unnecessary. It is perfectly acceptable (at least to me) to have the ligating virama by default produce the touching form and then upon using ZWJ in addition to produce the full ligature.

I therefore come to the conclusion that this model is the one suitable for the handling of Grantha virama forms while preserving the two important criteria: 1. semantic equivalence between all the forms, 2. allowing 99% of the users to get their desired virama form for each consonant by the simple single sequence C + Virama.

## Conclusion

Mr Ganesan has suggested that a chillu marker should be added to Grantha for some imagined semantic distinction it causes in Sanskrit. This is not true as has been variously proved by us. Mr Ganesan has made various false statements, including those regarding the use of danda-s and numerals in Grantha. He has even made the unnecessarily false statement that the font e-Grantamil by Elmar Kniprath is "widely used by Grantha publishing houses" while this is totally not true. This (the suggestion that "chillus" cause semantic distinction in Sanskrit) is merely one more of those false statements. While I have nothing personal against Mr Ganesan, I will not allow him to misrepresent the Grantha script, especially in such a serious thing as a Unicode proposal. I draw the attention of the UTC to the various examples I have provided proving semantic equivalence of ligated and unligated virama forms in Grantha, and also to the endorsement of this view by learned scholars of Tamil Nadu who have been publishing Grantha books for over twenty years. Therefore, while a ligating virama (again, NOT a chillu marker) is appropriate for Grantha, it should be decomposable to the regular virama and a chillu marker that is not decomposable this way should not be encoded for Grantha.

–o–o–o–