

Follow-up #2 to my Grantha proposal L2/09-372

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

2010-Jul-27

I submitted L2/10-085 titled “Feedback to Dr Anderson’s Grantha summary dt 2010-Jan-28” as both feedback to said Grantha summary and follow-up on my own Grantha proposal L2/09-372. The first “Miscellanea” section of that document was (mostly) feedback and the “Extended Tamil” section was follow-up. Now I discuss some more aspects that were left out of both those previous documents.

§1. Important Error: ‘Candrabindu’, not ‘Anunasika’

I have only recently noticed that in the context of the Grantha proposals, we have been using the name GRANTHA SIGN ANUNASIKA for the proposed 11301. Not only did Ganesan and I use this name this way, but even L2/10-167 which made recommendations carried this same name. However, the corresponding characters in the other Indic scripts – 0901, 0981, 0A81, 0B01, 0C01, 11080 – all use the word CANDRABINDU instead of ANUNASIKA. In fact, the word ANUNASIKA does not exist in the UnicodeData.txt file at all. Therefore this character should be renamed GRANTHA SIGN CANDRABINDU.

This said, I should note that in my separately submitted Extended Tamil proposal L2/10-256, I have asked for one character to carry the word ANUNASIKA in its name, but that is because a sign glyphically distinct from a candrabindu can be used for anunasika there. Since in Grantha it is only the candrabindu that is used for this purpose, the Grantha character should carry the word CANDRABINDU in accordance with the other Indic scripts.

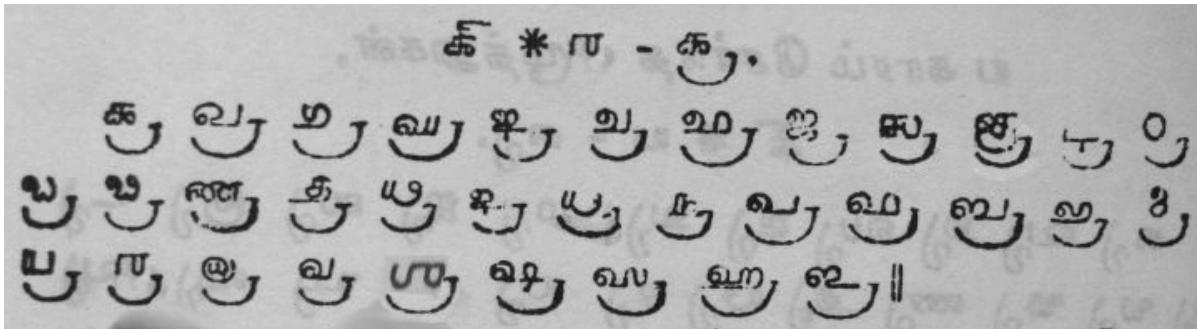
§2. Variation Selector Sequences

While the Grantha encoding is not complete, at least the number of core characters required for representing vowels and consonants has stabilized. Based on that, I feel confident enough to now formally request that along with the Grantha encoding, some variation selector sequences be provided to handle some glyph selection in Grantha where a choice is justifiably required and cannot be handled otherwise.

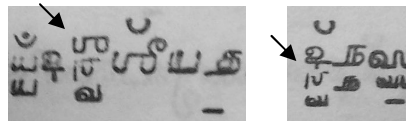
In §6.3.4 (p 39) of L2/09-372 (my Grantha proposal) I have mentioned a requirement for two VS sequences. I first recount them below and then further describe requirements for other VS sequences as well.

2.1. Sub-base RA instead of RA-vattu

In Grantha, as in many other Indic scripts, especially South Indian ones, consonant clusters are written as stacks, with the character representing successive consonants in a cluster being placed successively one below the other. In this, the consonant RA, when occurring (in non-initial position) in a cluster and represented in Unicode by the sequence VIRAMA + RA, is mostly rendered as a RA-vattu glyph, which is usually a curving stroke and seen (for Grantha of course) in the sample below from ref 1 p 24:



However, occasionally one comes across cases where the nominal form of RA is seen in sub-base position in stacks, such as in the following samples from ref 2, pp 56 and 375:



At first sight, it appears as if there might be a rule that says that the RA is represented by a RA-vattu only when present at the bottom of a stack, but upon inquiry, Vedic scholars using Grantha (some of whom are mentioned in pp 56, 57 of my proposal L2/09-372) opine that while it is most convenient to write the RA-vattu, it being almost a single stroke, the nominal form of RA placed below the base may be used in free variation anywhere in a stack. Thus, RA-vattu is optional in Grantha unlike in other scripts, and there is also no rule that it can only occur at the bottom of the stack. Therefore the RA-vattu and sub-base (form of nominal) RA are both equally C2-conjoining forms of the cluster non-initial RA.

Since these two forms are thus both equally valid, it would be appropriate to have a method to distinguish between the two. I propose that since the RA-vattu is the more common, both in Grantha and across Indic scripts, it be considered as the default, and a VS sequence be defined for selecting the sub-base form as follows:

VIRAMA + RA + VS1	→	sub-base RA
--------------------------	---	--------------------

The default sequence consisting of VIRAMA + RA would as usual either produce a ligature (as in SH·RA) or a RA-vattu (as in K·RA etc).

VIRAMA + RA

→

ligature or RA-vattu

To enforce the display of a RA-vattu even if a ligature is possible, the usual PR-37-prescribed sequence of ZWJ + VIRAMA is to be used.

ZWJ + VIRAMA + RA

→

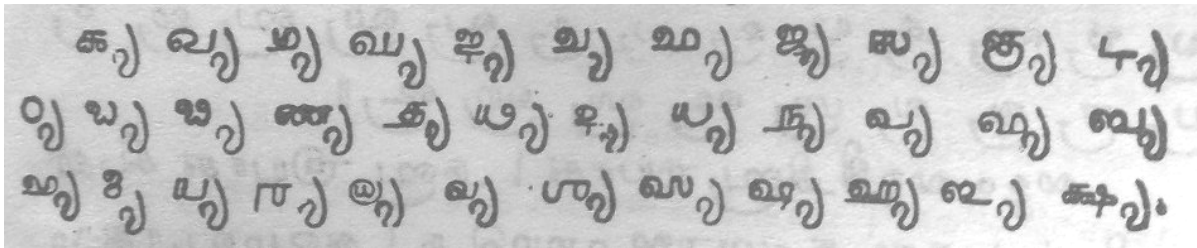
RA-vattu

It is to be noted that even though PR-37 only prescribed the ZWJ + VIRAMA sequence to generate the C2-conjoining form of a consonant, and does not discriminate between C2-conjoining forms, it would be ideal if in Grantha the sequence above produces only a RA-vattu and is not used for the sub-base RA because:

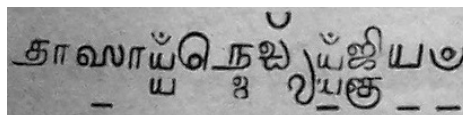
1. There are no two C2-conjoining forms of a single consonant, especially RA, in a any Indic script other than Grantha (to my knowledge).
2. Therefore PR-37 does not consider the problem of how to identify a particular (C1-or) C2-conjoining form from many.
3. Due to point 1 above, in all other Indic scripts, especially which have a RA-vattu, the sequence ZWJ + VIRAMA + RA is used for the selection of a RA-vattu only.

2.2. Sub-base YA instead of YA-phalaa

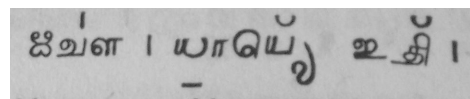
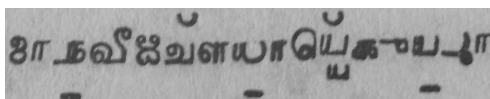
In Grantha, as in many other Indic scripts, a cluster-final YA is written in the form of a mark called a YA-phalaa. In Grantha this mark looks somewhat like U with a bend on the left and the right arm rising higher than the rest of the body, as seen in ref 1 p 23:



As an exception, in the cluster YA + VIRAMA + ANUNASIKA + YA, the cluster-final YA is always rendered as a YA-phalaa, as seen in ref 2 p 623:



Even apart from this exception, while the YA-phalaa is the default, sometimes sub-base YA may also be equivalently seen, as seen by the same Vedic text being written in both ways:



The above samples are from ref 2 p 177 and ref 4 p 250 respectively. This indicates that the YA-phalaa is optional and the sub-base YA may be equally used.

Therefore, while the YA-phalaa is the default, there should be a method to select sub-base YA as well. As discussed in the case of RA-vattu, both YA-phalaa and sub-base YA are C2-conjoining forms and therefore the sequence ZWJ + VIRAMA + YA cannot discriminate between the two. However, due to the widespread use of ZWJ + VIRAMA + YA in other Indic scripts to select the YA-phalaa, it is advised to have a VS sequence to select sub-base YA:

VIRAMA + YA + VS1	→	sub-base YA
--------------------------	---	--------------------

The default sequence consisting of VIRAMA + YA would as usual either produce a ligature (an archaic one such as T·YA, mentioned in my proposal in §6.3.1 p 37) or a YA-phalaa:

VIRAMA + YA (when cluster-final) → archaic ligature or YA-phalaa

To enforce the display of a YA-phalaa even if a ligature is possible, the usual PR-37-prescribed sequence of ZWJ + VIRAMA + YA is to be used.

ZWJ + VIRAMA + YA → YA-phalaa

2.3. Sub-base Tamil-style LA instead of sub-base Grantha-style LA

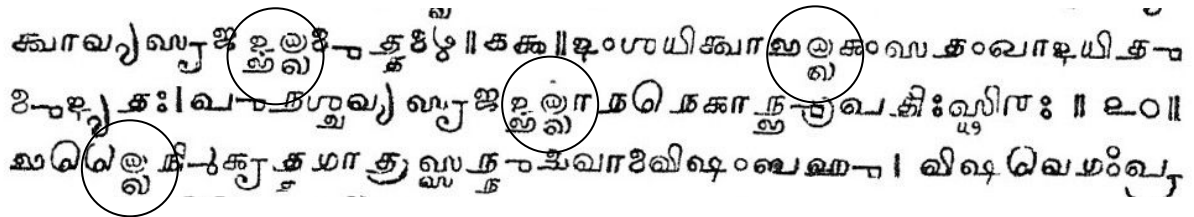
Apart from ligatures and special written forms such as the RA-vattu, repha and YA-phalaa, the chief method of representing consonant clusters in Grantha is as stacks, as discussed in detail in §5.4.2 and following sections in my proposal. As per the description given there, the consonant cluster LA + VIRAMA + LA would in Grantha be represented by a LA written below a LA, with any vowel signs applied to the stack as a single unit.

It goes without saying that both the LA-s are written as the same nominal Grantha-style glyph, since in Grantha unlike in other South Indian scripts like Kannada and Telugu the sub-base form of a consonant is the same as the nominal form of the consonant, albeit perhaps in reduced size. (The obvious exception is the RA-vattu.) The following samples from ref 5 p 445, ref 2 p 62 and ref 2 p 653 illustrate this:



However, very rarely but irrefutably one also come across printed matter in which the cluster LA + VIRAMA + LA is represented as a Grantha-style LA with a *Tamil-style* LA written below. That is, the sub-base LA is shown as a Tamil-style glyph. One sample from

ref 3 p 169 is shown below. Ref 3 uses both representations of LA + VIRAMA + LA equally and is hence replete with examples of both kinds.



Therefore I request that since both are valid attested forms of representing LA + VIRAMA + LA, and since, just as said for the RA-vattu and YA-phalaa, the sequence ZWJ + VIRAMA + LA cannot discriminate between either form of sub-base LA, I request that while the Grantha-style glyph is used as sub-base form of LA by default due to majority usage, a VS sequence be defined to select the Tamil-style sub-base LA, so:

VIRAMA + LA + VS1	→	Tamil-style sub-base LA
--------------------------	---	--------------------------------

The default sequence consisting of VIRAMA + LA would produce the Grantha-style sub-base LA (or a ligature if any such archaic ligature with LA is attested):

$$\text{VIRAMA + LA} \quad \rightarrow \quad \text{Grantha-style sub-base LA (/ lig)}$$

If in a particular font an archaic ligature with LA is the default, to enforce the display of a Grantha-style sub-base LA, the usual sequence of ZWJ + VIRAMA + LA is to be used as above:

$$\text{ZWJ + VIRAMA + LA} \quad \rightarrow \quad \text{Grantha-style sub-base LA}$$

It is noteworthy here that in Malayalam, the default representation of LA + VIRAMA + LA takes a Tamil-style LA in the sub-base position while the nominal Malayalam-style LA is at the top of the stack:

$$\text{ല് + ല = ല്ല}$$

This resemblance in behaviour is no doubt due to the evolutionary relationship between the scripts. (However, this similarity does not extend to the virama/chillu issue, as I have had occasion to discuss elsewhere.)

It is also to be noted that the VS sequence is not to be limited to LA + VIRAMA + LA because there are also other clusters such as SH + Virama + LA which exhibit this behaviour of using Tamil-style sub-base LA, as in the sample given below from ref 3 p 28:



It is to be noted that in Malayalam, the sub-base LA is always Tamil-style, even in SH·LA etc.

Therefore, the VS sequence being defined should not be limited to L·LA and hence the generic sequence of VIRAMA + LA + VS1 should be appropriate. It would also enable the display in isolation, if desired, of this Tamil-style sub-base LA, just as the sequence ZWJ + VIRAMA + LA would show the Grantha-style sub-base LA, when used with NBSP.

2.4. Three different virama forms

The chillu/virama issue has been debated in the context of Grantha for over a year now. In fact, to those who properly understand the Sanskrit language, the Grantha script and their mutual relationship, there is no confusion in this issue at all. In my proposal L2/09-372 I have shown evidence from printed material for the existence of three different forms of virama which are glyphically and/or behaviourally distinct from one another but are all however mutually semantically equal. Numerous traditional Vedic/Sanskrit scholars of Tamil Nadu, whose names and qualifications are given in L2/09-372 p 56 and L2/10-233 have endorsed this analysis of Grantha. Therefore the Unicode encoding of Grantha should represent this aspect of Grantha carefully and correctly.

Hitherto, I have been advocating the usage of combinations involving ZWNJ, ZWJ and a separate virama character called a LIGATING VIRAMA to handle the situation. However, the joiner situation in Indic scripts, especially with regard to the far South Indian scripts which exhibit more than one form of representing isolated vowelless consonants (referring to the Malayalam chillus), is already complicated. This has given room for others to find fault with my previously mentioned LIGATING VIRAMA solution that it confounds the joiner situation even more.

That solution also asks that the LIGATING VIRAMA be canonically decomposable to the regular virama of Grantha. Till now, canonical decomposition is only provided in Unicode when there are different but glyphically (and semantically) equivalent character sequences for composing the same text. While the LIGATING VIRAMA would be semantically equivalent to the regular virama, it has no glyphic form, being merely a technical device like 17D2 KHMER SIGN COENG. Therefore the decomposition might be not entirely self-evident. This is one more reason why this LIGATING VIRAMA solution may not be entirely satisfactory.

Therefore I have suggested in my separate document L2/10-___ “Comments on some more documents of Naga Ganesan”, the alternative solution that *if there is a need to provide unambiguous methods of selecting one of the three virama forms possible for a consonant in Grantha*, VS sequences should be defined for that purpose. Three VS characters VS1, VS2

and VS3 should be used, and the sequences of GRANTHA VIRAMA + VS1/VS2/VS3 should be allotted to consistently display one virama form each.

This solution is better than the LIGATING VIRAMA solution in that:

- 1) It avoids joiners altogether and hence does not complicate any existing situation concerning them.
- 2) It avoids the encoding of an additional character called a LIGATING VIRAMA which has no real appearance and whose purpose may not be clear to all.
- 3) It hence also avoids the need for canonically decomposing that character to the regular virama, which may not be considered entirely appropriate.
- 4) It highlights the fact that the difference between the various virama forms is a mere glyphic one (which the VS characters are intended to represent) and that there is no semantic difference.
- 5) It enables searches for Grantha text written with one virama form to return equivalent text written with another virama form, and words carrying one form to be collated equivalently to words carrying another form, since the VS characters are ignored (or at least can be ignored) in such text operations due to their UCA weights being zero, thereby ensuring the semantic equality of the virama forms.

Thus I request, *if there is a need to provide unambiguous methods of selecting one of the three virama forms possible for a consonant in Grantha*, appropriate VS sequences are defined for the purpose and no foreign mechanisms such as a chillu marker are introduced.

§3. Old-style vowel signs for Vocalic L and LL

In my Grantha proposal (L2/09-372 §3.5.1) I have demonstrated the existence of distinct old-style vowel sign forms for the vowels Vocalic L and LL. These forms involve the writing of the appropriate glyph below the consonant the sign is applied to, unlike the default in Grantha wherein the glyph is written to the right of the consonant. In all other Indian scripts, the vowel signs for Vocalic L and LL are *always* written below the consonant, and it is possible that this was so previously in Grantha as well. However, the fact is irrefutable that in modern written and printed Grantha the vowel sign is placed to the right. Thus, since both ways of writing the vowel sign are attested, both are valid.

In my proposal I have asked for these two old-style vowel sign forms to be encoded as separate characters from the default forms but with a canonical decomposition to the

corresponding default forms in order to maintain semantic equivalence. I am informed that the matter of separately encoding these characters has been tabled for future discussion. Therefore I would like to here record some more thoughts about these characters.

Till now, as far as I know and as I have mentioned above, decomposition is only provided in Unicode when there are different glyphically (and semantically) equivalent character sequences for the same text. While the old-style forms in question are indeed semantically equivalent to the new-style forms, they can not be said to be glyphically identical due to the matter of their position of application to the consonant being different.

Therefore it is possible that the UTC may not favour the requested canonical decomposition. In such case, the semantic equivalence for the purpose of text search and collation may be handled by “higher level protocols”. I have already remarked in my proposal (L2/09-372 §3.5.2) that the vowel sign AU in Grantha is written (as in Malayalam) in two ways, as a single sign on the right or as a two-part sign. To achieve semantic equivalence between these two forms in text search and collation, higher level protocols are certainly required since canonical decomposition is not possible here. A similar mechanism may be employed for the present matter of the vowel signs also.

Since the same Grantha text may be found to show both ways of writing the vowel sign, there should be a way of representing this distinction in encoding. As far as I can see the only way is to encode two separate characters for the old-style forms. Therefore these two separate characters should certainly be encoded. Decomposition however may be avoided by relegating the task of preserving the semantic identity to higher level protocols.

§4. References

1. Grantha Bāla Pāṭhamālā, Heritage India Education Trust, Chennai, India, 2003.
2. Taittirīya Saṃhitā, Heritage India Education Trust, Chennai, India, 1980.
3. Śrī Hālāsya Māhātmyam, http://ambaa.org/pdf/halasya_mahatmyam.pdf
4. Taittirīya Saṃhitā Pada Pāṭha, Mylapore Veda Adhyayana Sabha, Chennai, 1996.
5. Taittirīya Brāhmaṇam, Heritage India Education Trust, Chennai, India, 1980.

-o-o-o-