Comments on Mr Ganesan's Grantha Proposal

Shriramana Sharma

2009-Aug-15

Indian Independence Day

SUMMARY:

1) A separate chillu marker should not be encoded since chillus in Grantha carry no semantic difference against virama forms. There are other logical methods of contrasting chillus and virama forms to cater to users' writing preferences.

2) The danda-s from the Devanagari block and the numerals from the Tamil block should not be disunified as Modern Grantha's danda-s or numerals are not glyphically or behaviourally distinct.

3) No attestation has been given for a Grantha OM symbol and hence it should not be encoded now.

4) It has not been clearly stated that glyphs proposed for short E and O are not attested for Grantha but are newly suggested based on old forms of writing Tamil.

5) Tamil LLLA, RRA and NNNA should not be disunified but used as-is from the Tamil block.

I. OBJECTION TO ENCODING A SEPARATE CHILLU MARKER

The false examples given to supposedly prove the semantic significance of chillus

In page 5 of his proposal, where Mr Ganesan purports to show a semantic change caused by the usage of a chillu marker, he has not explained the meaning of any of the words at all. As it is unlikely that all Unicode people can understand Sanskrit, one would think that anyone who would like to show a difference in meaning would write it in English, as Eric Muller has done in N3126 where he showed that വൻയവണിക means big curtain and വന്യവണിക means wild forest in Malayalam.

Further, if anyone wants to show contrast between two usages, one must keep everything else constant and make only the single change in writing and show how it changes the meaning. Eric Muller has correctly done this in N3126 in the case of വൻയവണിക and വന്യവണിക where the only difference is in the usage of the chillu and all preceding and following sequences are the same.

In his proposal, all Mr Ganesan has done is to write one word with a chillu and write another different word without a chillu and he claims that a difference in meaning has been created! (This is all the more preposterous seeing that he has not explained any of the meanings at all!)

The meanings of the word he has written are: സന്മാഗ്ഥും – "the righteous path", accusative case; ജന്ഥം – "birth"; ഥഹവകക്ഥായും – "a deed in service of the Lord"; உ௳ட – "intense".

It should be noted that the proper Sanskrit word is ജന്ഥ, not ജന്ഥം. Further, Sanskrit does not permit an anusvāra in final position except when a consonant at the head of another word follows. So it is poor form to write an anusvāra at the end of free words (though lots of texts do this for some imagined convenience). The proper way is to write either the chillu form of MA ൿ or MA with an

overt virama ். Above, we have written the words exactly as Mr Ganesan did only to quote them as-is, and this does not mean that we endorse writing them so. We employ the correct usage below.

Anyway, Mr Ganesan has *not* shown what change in meaning has been created by the usage of the chillu as against the usage of a consonant with explicit virama / consonant ligature / stacked consonant cluster. He cannot show it simply because it does not exist. He is perhaps ignorant of the fact that even if the words had been written ஸநாஹ௲ and ஹஹவதாயு௲ without the chillu-s, or ஜநூஉ and உசகூட with chillu-s, it would mean the same.

Chillu-s are not used only at the ends of words

Perhaps Mr Ganesan is labouring under the misapprehension that chillu-s indicate the end of words whereas stacked or combining consonants mean a continuing word. Modern printed texts show that this is not true for Malayalam, and we will show that this certainly not true for Sanskrit.

The photo below is from page 760 of ref 1. Here you see a chillu in the middle of a word. The corresponding Devanagari version at page 19 (penultimate line under 3-14-2) of ref 2 shows that this is a single word समिन्थसे = sam + inthse where sam is a prefix attached to verbal roots and inthse is present tense second person singular of the root 'indh', to kindle.
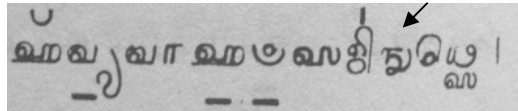

Further see the following picture from page 15 of the Ārcika part of ref 3.
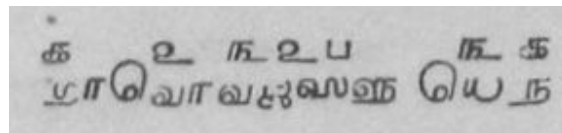
Compare it to the corresponding Devanagari text वत्सं न at the top of page 22 of ref 4. vatsa = calf (of a cow, usually). See also Ralph Griffith's translation of the Sama Veda: http://www.sacred-texts.com/hin/sv.htm and scroll down to Book II, Chapter II, Decade I: "Like milch-kine lowing to their calves!". Thus it is confirmed that vatsam is a single word yet it has a chillu in the middle.

Therefore in Sanskrit it cannot be concluded seeing a chillu that it is the end of a word.

Further see the following sample from page 750 of ref 1:
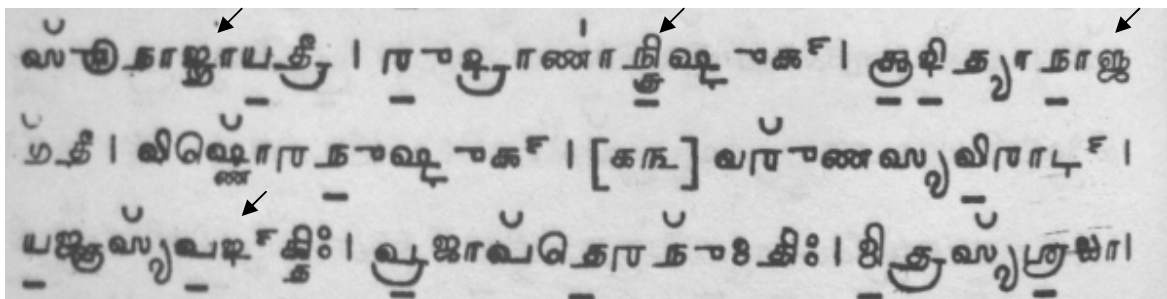
The arrows on the second line indicate consonant sequences which have formed stacked forms or ligatures. That these are in fact joining points of two different words is obvious when you compare the Devanagari versions वसूनां गायत्री, रुद्राणां त्रिष्टुक्, आदित्यानां जगती (top of page 7 of ref 2). Therefore it cannot be concluded seeing a consonant ligature or stacked form that it is part of a single word.

Also see the arrow-marked part in the bottom of the same image showing how the single word pankti is written using an explicit virama in the middle and compare the Devanagari पङ्क्ति.

Further, the image on the right from p 90 of ref 1 shows that NA, which can form a chillu form, does not form one even at the end of a word, but joins with the following VA in the sequence yasmin + vāyuḥ (two well known simple Sanskrit words meaning "in which" and "air") to form a consonant ligature.
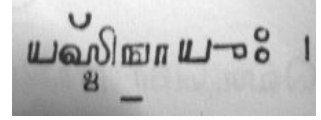
Thus there are absolutely no orthographic rules for Sanskrit, whether written in Devanagari or Grantha, as to where ligatures, stacks, explicit virama forms or chillu forms must be used in consonant clusters. It is a fact well known to all Sanskritists that it is often possible to properly identify word boundaries in written or even spoken words only when one is knowledgeable in the Sanskrit language, and there are no orthographic rules which will help here. Mr Ganesan's imagination that chillu-s change meanings will not alter the true nature of Sanskrit that they do not.

Other false cases of chillu-s allegedly causing semantic difference

Mr Ganesan goes on to claim (and 'is lost' should be appended to complete his sentence) that:

There are many words where the meaning is lost if the contrast between chillu consonant and consonant conjunct with vertical stacking of consonants or consonant with explicit virama (produced using ZWNJ).

This is totally preposterous! Mr Ganesan has not provided a single proof of this. Let him show one real example of this behaviour. He cannot, simply because none exists. Mr Ganesan's insufficient experience with the Sanskrit language and orthography is crystal-clear to Sanskrit scholars like me.

Mr Ganesan continues:

There are whole books (e.g., Raaghava PaaNDaviiyam) written with different chillu consonants occurring in the middle of words. If the chillus are converted to something else (e.g., stacked conjunct) the meaning will change producing a punning effect.

Let Mr Ganesan provide us samples from those books and we will analyse them for him and the UTC. It is a well known fact that śleṣa (double meaning) is a commonly used wordplay in Sanskrit literature. It is quite possible that such differences occur when letters are split one way or the other. Even in the well known words of Kalidāsa's Raghuvaṃśa: "जगतः पितरौ वन्दे पार्वतीपरमेश्वरौ", splitting the final word as पार्वती + परमेश्वरौ would refer to the divine couple Parvatī and Śiva, whereas splitting it as पार्वतीप + रमेश्वरौ would refer to Śiva and Viṣṇu. This happens in all languages and scripts.

But how does this relate to the present case? We have been asking what semantic difference is caused by the usage of the chillu, not caused by splitting Sanskrit words one way or the other.

If a writer wishes to show a consonant separately, in order to clarify it from the next word, he can always insert a ZWNJ or a space. See for example the following shloka (4-21) of the Bhagavad Gita (available for download from http://sanskritweb.net/sansdocs/index.html#GITABIG) which ends:

कुर्वन्नाप्नोति किल्बिषम् ।

Here the proper splitting of words is: कुर्वन् न आप्नोति किल्बिषम् । It means "कुर्वन् = he who does so, न = does not, आप्नोति = incur, किल्बिषम् = sin". (Anyone wanting to check please see the translation of the 21ˢᵗ verse under http://www.dlshq.org/download/bgita.htm#_VPID_13.) If one were to split it as कुर्वन् आप्नोति किल्बिषम् (which is also a valid possibility under Sanskrit sandhi rules) it would mean "he who does so will incur sin". Therefore one may want to write the original text as: कुर्वन् नाप्नोति किल्बिषम् which would clarify that the extra NA is not a product of the sandhi between कुर्वन् and आप्नोति.

If such a text were to be written in Grantha, the word कुर्वन् would be terminated by a chillu OR a NA with explicit virama. While we will not deny that the preference is for writing a chillu, we point out that even using the explicit virama will not change the meaning here.

கூர்வுந்ᵣ நாவ்ொ்ாதி = கூ்வுந்நு நாவ்ொ்ாதி = கூ்வு்ணாவ்ொ்ாதி

As it is, it is only the visible splitting of the written words that has *clarified*, not *changed*, the intended meaning, and the chillu has nothing to do with it, since the same clarification is achieved by displaying the virama form too. Thus the chillu is in no way semantically special.

Sanskrit is independent of scripts and does not give special value to chillus

Mr Ganesan continues:

> *Often in manuscripts there is no gap between words in the Grantha script, and a unique chillu-consonant encoding is essential for proper representation of plain-text data. For example, in the famous Bhagavat-giitaa, different commentaries change the meaning depending on the chillu presence or absence, to bring out various philosophical points.*

Mr Ganesan perhaps does not know that it is only the splitting of the word, such as in the example from the self-same Bhagavad Gita which we gave above, that causes the difference in meaning, and not the usage of the chillu. As pointed out above, the same would be achieved by writing a consonant with overt virama. And as for different commentaries changing the meaning depending on the chillu, this is simply laughable. Does Mr Ganesan think that all the commentators used only Grantha manuscripts? There have been many commentaries on the Bhagavad Gita by North Indians. Are we to believe that they came all the way to Tamil Nadu to get a Grantha manuscript and check whether there is a chillu in a particular place or not just to determine the meaning?

The Sanskrit language as a whole has been all along script-independent. As the centuries passed, writers evolved various scripts and perhaps various techniques in the various scripts to denote various aspects of the normal and the Vedic Sanskrit language, but the language itself, and

people who know the language like the commentators on the Bhagavad Gita to which Mr Ganesan refers, do not depend on orthographic features such as chillu-s to determine the meaning of the holy texts, unlike the situation in Malayalam where even people who know the language might read a different meaning depending on whether a chillu or virama was used.

Again, we believe that Mr Ganesan is writing like this (that the meaning will change depending on the presence or absence of a chillu) because he thinks that chillus occur only wherever a dead consonant occurs at the end of a word. The examples we have provided above from actual authentic Vedic texts printed in Grantha prove that he is seriously mistaken.

This being the situation, we are dismayed at Mr Ganesan's continuing to submit such imperfectly argued proposals to the UTC with this so-called "revised" version, of which he never sent a copy to us despite our repeated requests to him. Nor had he first sent his original proposal when requested. When we have asked him a question as to the chillu-s, how it is that he does not answer us directly if he is confident of the validity of his stand? How is it that he instead speaks directly to the UTC who are certainly experts in computer science, typography and scripts but not the Sanskrit language? We hope that Mr Ganesan has no ideas of passing his proposal without the approval of the Sanskrit scholars of Tamil Nadu who even today use the Grantha script.

Grantha chillu-s and Malayalam chillu-s are not similar in behaviour or nature

Now before continuing our commentary on Mr Ganesan's proposal, we wish to state the following: Whatever was discussed in the whole of the cyberworld and real world in favour of encoding atomic chillus for Malayalam can apply to Malayalam and Malayalam alone. How does it follow that the same arguments will apply to Grantha too? We will now prove that the arguments provided in support of encoding separate chillus (or a chillu marker) do not apply to Grantha.

In justifying the separate encoding of chillaksharams in Malayalam, Eric Muller, the author of N3126 says, and we quote, with explanatory words in parantheses and separation into numbered paragraphs alone ours:

> *Without denying the relationship between a cillaksaram and its underlying consonant, the examples above show:*
>
> 1) *that from the graphic point of view, they (a chillaksharam and the chandrakala form of its underlying consonant) are used constrastively (sic) (in the sense that there are attested cases where one meaning is construed when the chillaksharam is used and another is construed when the chandrakala form is used), and*
>
> 2) *that the current representation (wherein a chillu must be written as a consonant + virama + ZWJ) makes the joiners semantically significant.*
>
> *Because of:*
>
> 3) *the general principle on joiners (that they should not be semantically significant and should only used to denote rendering preferences),*

4) and the application of this principles in environments such as IDNs (where joiners have been officially declared as ignorable, and it therefore being impossible to register a website വൻയവനിക.com to sell big curtains as different from വന്യവനിക.com describing a forest resort),

the current representation is problematic.

5) In addition, there is uncertainty on the consonant underlying some cillaksarams (creating the IDN security problem where it is is possible for a malicious party, when there is a website മലർ.com encoded with RA+virama+ZWJ, to register a parallel website മലർ.com encoded with RRA+virama+ZWJ and spoof people into going to that website instead).

These are the only reasons provided for encoding separate chillaksharams in Malayalam. Let us now examine the case of Grantha:

1. In the case of Grantha, the chillaksharam and the overt virama form of a consonant are *not* used contrastively, despite Mr Ganesan's baseless claims to the contrary. They are in fact used interchangeably. This is proved by the following photos:
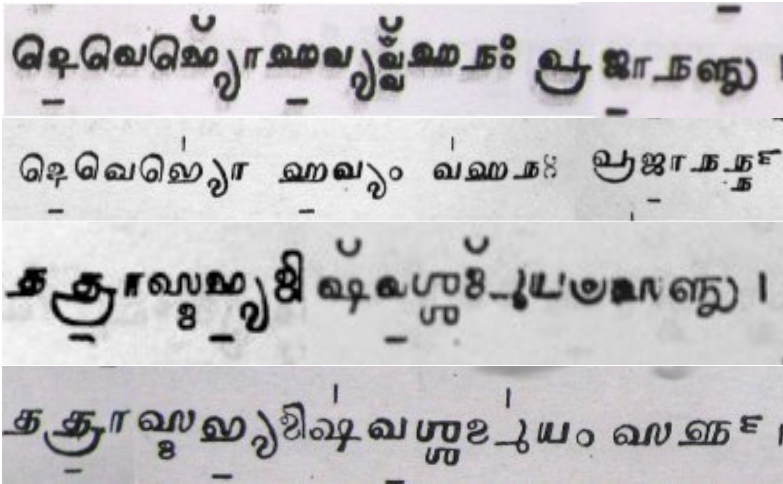
These photos, being in order from p 285 of ref 1, p 36 of ref 9, p 449 of ref 8 and p 72 of ref 9, show how the chillu of the double NA (NA + virama + NA) is used interchangeably with the respective virama form. The text in the first pair of images 'devebhyo havyaṃ vaha naḥ prajānann' means 'O thou knowledgeable One, take our oblations unto the gods!'. The text in the second pair of images 'tatrāsmabhyam iṣavaśśarma yaṃsann' means: and I quote Ralph Griffith's translation (http://www.sacred-texts.com/hin/rigveda/rv06075.htm, verse 11) "there may the Arrows shelter and protect us". There is no difference in meaning here and the different style of writing is merely due to the different source text (ref 1 vs ref 9, ref 8 vs ref 9).

The following image (comprising 6 small photos stitched together) from pp 27, 28 of ref 11 (an introductory text for those learning Grantha) shows how the chillu-s of NA and MA are used interchangeably in equivalent contexts:
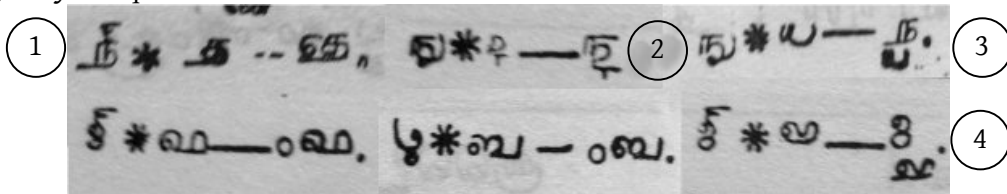
This also illustrates how usage of chillu is equivalent to usage of consonant ligature (2), stacked consonants (3), and the overt virama is also equivalent likewise (1)(4). This further underlines the identity of the chillu and overt virama forms, while disproving Mr Ganesan's claims that usage of chillu forms as against ligatures or stacked forms or virama forms causes difference of meaning.

In any case, it is the burden of those who advocate a separate chillu marker to show proofs of non-interchangeability of chillu-s with virama forms, like മനുഷ്യൻ vs മനുഷ്യന് കൊടുക്കുന്നു for Malayalam. However, that is impossible here since interchangeability is the fact in Sanskrit.

In summary, however he may try, now Mr Ganesan cannot claim that there is a semantic contrast between the graphical usage of chillu-s as against overt virama-s. Now returning to the other points stated by Eric Muller in support of Malayalam chillu-s:

2. As there is no contrast between the chillu-s and overt virama-s, we do not use ZWJ (or ZWNJ) to encode chillu-s. Therefore there is no semantic significance for any joiners.

3. The general principle that joiners should carry no semantic significance is not broken.

4. Since there is no semantic difference between using a chillu and virama, a native Grantha user would not expect to register �സകൃഴ.com as distinct from സകൃൾ.com.

5. As no two Granthaksharams have the same chillu form, there is no possibility that anyone can spoof a website whose name is in Grantha exploiting such a weakness.

Therefore none of the reasons provided for the separate encoding of Malayalam chillaksharams can be applied for the encoding of chillaksharams or chillu marker in Grantha.

Why a separate Grantha chillu marker should not be encoded

If a separate chillu marker is encoded despite this, people are liable to think that the chillu has some special significance, which it does not. The UTC has a huge social responsibility in that millions of people trust in what is published by them. If they publish a chillu marker in accordance with Mr Ganesan's wishes, they would be doing something which would confuse a lot of people and give them the wrong impression about the Grantha script.

To exemplify this, I would like to briefly recount here an experience of mine. Mr Ulrich Stiehl of Germany, the owner of http://sanskritweb.net asked me to write for his website an article on Vedic Svara-s. (It is even now available at http://sanskritweb.net/sansdocs/index.html#IPA). In it I used the marker 0951 for Svarita, as is customary in most parts of India. However, he demanded to know how I could do that when Unicode labeled 0951 as DEVANAGARI STRESS SIGN UDATTA. Upon this, I contacted Unicode with a complaint after which the explanatory note about it mostly being used for Svarita was added. I do not know how many other people thought the single upper stroke always meant the Udatta until that explanatory note was added and how many people saw that note.

Why should something imperfect be done only to later have a half-amendment done to it? If a separate chillu marker is encoded one would have to explain why it requires a separate encoding (which is impossible, as is being explained by me with so much effort), and then add a note that that it applies to only some of the consonants, not all of them.

The hypothetical case of using Grantha for transcription of Malayalam

Now we shall continue with our examination. Mr Ganesan continues:

Also, there are words like kaNvalayam and vanyavanikaa that can be split(?) with or without a chillu depending on whether a unit taken as Dravidian word etymologically. These family of words have been discussed in great detail during chillu encoding in UTC documents, and to represent these kinds of double-meaning words and Malayalam web pages in one-to-one transliteration, Grantha chillu marker sign encoding is necessary

We feel that Mr Ganesan is attaching too much importance to using and extending Grantha to denote other languages than Sanskrit. Though it is a good thing to extend Grantha, too much of anything is...

Now let us assume, though it is not true and we shall so prove later on, that it is necessary to encode a chillu marker to represent Malayalam in Grantha script. The question now arises whether it is *worth it* to encode a chillu marker in Grantha just for the purpose of representing Malayalam.

We agree that enabling the transliteration in Grantha of other languages than Sanskrit is good in itself, and will help the popularization of the script and so we accept it when, for example, Mr Ganesan suggests the inclusion of the symbols for short vowels E and O. These is acceptable because it changes neither the current behaviour of the script nor the relation of the script with its natural language, Sanskrit. The demarcation line is, however, that no change should be made to the script that would damage the glyphic or behavioural nature of the script as it exists, or the way that the script is used to denote its natural language.

Would Mr Ganesan suggest a chillu marker for transliterating Malayalam for Devanagari too?

As we have explained above, if a chillu marker is encoded separately for Grantha, people are liable to think that it is something special and has a distinct semantic significance. So wherever there are Grantha chillu-s, even when the text is in Sanskrit and not the hypothetical Malayalam, people will start "seeing things" and imagining differences in meaning which are not there. This would alter the way that Grantha script has been representing the Sanskrit language till today. This would also hide the behavioural nature of the script, especially the ligating behaviour of C + virama to form chillus, since only C + chillu marker would be rendered as a chillu.

[It cannot be allowed, if a chillu marker is encoded, that C + virama is also rendered as a chillu as this would result in visual ambiguity between C + virama and C + chillu marker, creating a security problem such as described in point 5 of the justification in atomic encoding of Malayalam chillu-s.]

As to what semantic significance the chillu has, Mr Ganesan has not said yet and he can never say, after we have demonstrated that virama and chillu forms are used interchangeably.

Therefore adding a chillu marker just for the hypothetical need in distinguishing Malayalam sequences written in Grantha is not advisable. Further, seeing as even Grünendahl does not show any chillu-s in Grantha for LA and LLA in page 16 of ref 5, what is Mr Ganesan going to do for that? Is he going to further *invent* chillu-s for those two consonants? And this is assuming that he would use the RA chillu for RRA too.

There should be limits to what Mr Ganesan can do to the Grantha script to further his goal of popularizing Grantha. One must be pragmatic, and there are some unspoken rules regarding graphological tradition of the Grantha script or any other script developed over the centuries which any person respecting that script must follow, some limits which one must not cross. The Grantha script is not a playground for anyone. Mr Ganesan should not attempt to make a Universal Transliteration Script of Grantha. That is an impossible task, even considering only Indic scripts. Each script has its own idiosyncrasies which can never be correctly represented in other scripts.

Returning to the present matter, it is clear that the Grantha script does not have sufficient glyphic resources to denote Malayalam's chillu-s, *even if* Mr Ganesan's chillu marker is encoded.

Mr Ganesan's proposal also begs the question of how the combination of C + chillu marker, where C does not have a chillu form is to be rendered. If it is rendered with a virama, it is again a phisher's birthday present and serious security problem because there would then be visual ambiguity between the sequences C + virama and C + chillu marker which is a security problem as described before. A proper font should render it with a dotted circle, but we doubt that the UTC will entrust the security of Unicode's users to unknown font makers. If the font maker, due to whatever reason, falls back to a virama, an unsuspecting user may fall prey to such tactics of malicious parties.

Suggested method to achieve contrastive display of chillu vs overt virama form

Now we shall show that, if at all it is necessary, it is possible to give a proper transliteration of Malayalam consonants without using any separate chillu marker.

We have repeatedly proposed that where C is a consonant with a chillu form, and such chillu form is available in a Grantha font, C + virama should be by default rendered as that chillu form. Otherwise, i.e. where C does not have a chillu form, or such chillu form is not available in a font, C + virama will be rendered as C with an overt virama. There are no IDN security issues associated with this fallback method since we never encoded a separate chillu marker which would appear the same as virama in case of fallback thereby giving scope for malicious activity.

Further, if one wants to display the form of C with overt virama though C has a chillu form, the ZWJ can be placed in between C and virama, so: C + ZWJ + virama. This is in similarity to the usage of ZWJ prescribed in the Unicode paper ref 6 wherein it both joins and separates consonants in a consonant cluster – joins them because an overt halant is not shown, separates them because a ligature is not shown. Similarly, here the ZWJ joins the consonant and virama since the virama is a combining mark and as such should be applied to its base consonant, and separates them so that they do not form a ligature. Since ref 6 prescribes C + ZWJ + virama + C2 to be rendered as the full form of C with the sub- or post-base form of C2, if the overt virama form of C is desired to be followed by the full form of C2, C + ZWJ + virama + ZWNJ + C2 is to be used. If one wants the chillu form of C to be followed by the full form of C2, C + virama + ZWNJ + C2 is to be used.

Now armed with these rules, we proceed to show how the desired contrasts in the case of Malayalam chillaksharams can be achieved without the use of a separate chillu marker in Grantha:

ஆ உஞூஷ்ரு்ங கொட꧀ஊசூரூஊௗ   ஆ உஞூஷ்ரீ் கொட꧀ஊசூரூஊௗ

(Please read the long O vowel sign as short O, since my font currently does not support the proposed short vowel sign.) On the LHS, we have at the end of the word "manuṣyan", the chillu represented by NA + VIRAMA, and on the RHS we have the overt virama form (Grantha equivalent of Malayalam chandrakala) represented by NA + ZWJ + VIRAMA.

The objection can be made that this method depends on the font to provide a chillu glyph for NA + VIRAMA, in the absence of which the overt virama form will be displayed as a fallback and this will cause the loss of ability to retain the exact form as shown in Malayalam. To this, we will reply that even if a separate chillu marker is encoded, it is up to the font maker to provide the actual chillu glyph for NA in the font and for the sequence NA + chillu marker to be rendered as that glyph, and if such a separate chillu is not provided, fallback will certainly happen. So this problem is not unique to our side but exists on Mr Ganesan's side as well.

And now for the big curtain / wild forest problem:

வந்யவநிக   வநுவநிக

In accordance with the principles we stated above, the LHS will be encoded by VA + NA + virama + ZWNJ + YA, and the RHS by the same without the ZWNJ. This usage of ZWNJ is in accordance with the rules outlined in page 14 of the Unicode paper ref 6 wherein it states:

> *For all Indic scripts, ZWNJ can be used in a sequence < C1, virama, ZWNJ, C2 > to*
> *explicitly restrict the display to the level-3 alternative, the overt halant form.*

Only, here the chillu replaces the overt halant form, since that document did not consider the problem of chillaksharams. The rule intends to state that ZWNJ in the middle of a consonant cluster halts any conjoining behaviour between the sections before and after it and thus effectively the two units of a consonant cluster separated by a ZWNJ must be rendered as if they were independent. Thus since we define the default rendering behaviour of C + virama to show the chillu, we can achieve the desired display using only a single ZWNJ. Note that here the ZWNJ is not part of the sequence that is needed to encode the chillu. It is only for separating the NA + virama sequence from the YA following it, after which that sequence is automatically rendered as a chillu if available.

Of course, if it were desired to display (for whatever reason):

வநீ்யவநிக

where it is desired to both prevent C + virama conjoining with the following consonant as well as prevent C being displayed as a chillu, the sequence C + ZWJ + virama + ZWNJ + YA can be used as said hereinbefore.

Refutal of objections to the above method

It may here be objected that our proposal makes the ZWJ semantically significant when Grantha is used for transliterating Malayalam, since the ZWJ in உஞூஷ்ரீ் (NA + ZWJ + VIRAMA) causes change of meaning. We believe that this objection is meaningless. Grantha is proposed for the

purpose of encoding Sanskrit, not for Malayalam. Malayalam support is an extra, not an essential. And as an extra, it can be provided only to the extent that it does not affect the relation to Sanskrit.

The fact remains that in Grantha script denoting Sanskrit chillu-s and overt virama forms are used interchangeably. This may not be the fact in Malayalam script denoting Malayalam. But that is not sufficient justification to alter the nature of the Grantha script stating a hypothetical need to be able to transliterate Malayalam exactly. It is also not appropriate for Mr Ganesan to suggest a script reform for transliteration purposes. A script reform is made only when it is desired to improve the way a script is used to denote its native language(s), not how it denotes other languages.

The only remaining possible objections are that our method makes it:

1. impossible to register வந்யவதிக௯.com as distinct from வந�வதிக௯.com since IDN ignores ZWNJ

2. possible for உஐ௴ீ.com using RA+virama to be spoofed by உஐ௴ீ.com using RRA+virama (please read that ௴ீ as RA-chillu which my font does not currently support).

Here the second objection arises only if Mr Ganesan were to suggest that we should render RRA + virama in Grantha by the same chillu as for RA since there is no chillu for RRA in Grantha and in pursuance of the imagined ideal of one-one equivalence with Malayalam (where the chillu for RA and RRA is the same glyph). Inventing a new chillu for RRA for Grantha is also unacceptable.

Now we hardly think that anyone is going to want to register a Malayalam name for a domain encoded in the Grantha script! This is simply too far-fetched to merit any serious consideration. The only valid case that can be made out for transliteration is for the *content* of websites, not the *names* of websites. If at all the names of websites should also be transliterated, it should be only in the address bar and viewpane of a web browser and not in the IDN registry. Such a preposterous suggestion as to there being a need to register a Malayalam domain name written in Grantha will be rejected by the most simple-minded of people, what to say of the well-learned experts in the UTC.

Therefore we do not believe that this can be considered sufficient reason to encode a separate chillu character for Grantha, whose potential undesired side-effects are listed by us above in terms of people misunderstanding the nature of the Grantha script and misreading Sanskrit texts written in the Grantha script wherein chillus and overt viramas occur in various combinations.

Thus we submit, after the very long dissertation above, that a separate chillu marker should *not* be encoded for Grantha. If anyone wishes to encode it anyway, they should give answers to all the arguments presented in these pages. Otherwise, it would only mean that they have bid all intellectual honesty goodbye. We believe that the intellectually honest UTC will see the truth in our argument and not approve Mr Ganesan's request on this score (of encoding a separate chillu marker), seeing as he has not even shown what semantic contrast is actually caused by the usage of a chillu vs an overt virama marker except for claiming that such contrast is created.

Thus a separate chillu marker should not be encoded. The display of chillus is up to the font developers. A mechanism is also provided to specifically request display of overt virama-s.

II. OBJECTION TO ENCODING SCRIPT-SPECIFIC DANDA-S

In connection with the matter of encoding script-specific danda-s, Mr Ganesan says:

UTC document ISO/IEC JTC 1/SC 2/WG 2 N3452, Section F.7.2. (For new scripts proposed for encoding) states that "The existence of the use of Dandas in orthographies for a script proposed for encoding is generally taken as sufficient justification for encoding of script-specific Dandas for that script."

We do not disagree, only we would prefer to go along with the existing behaviour of all Indian scripts using the generic Danda-s encoded as 0964 and 0965. But Mr Ganesan continues:

Grantha is a newly proposed script in Unicode and the justification exists for providing Grantha-specific danda signs in Unicode encoding. Danda signs are NOT a recent borrowal from Devanagari signs as Indologists affirm the danda presence in palm-leaf manuscripts.

We believe that Mr Ganesan has misconstrued the import of the P&P document. When the P&P document says that the danda-s from the Devanagari block are used for the other Indian scripts, they never meant to say that danda-s used in non-Devanagari scripts are a borrowal from the Devanagari script. They only mean that since there is no point in disunifying the danda-s *which have already been placed in the Devanagari coderange* the same danda-s are used for the other scripts. This is a question of whether to disunify or not, not which script used danda-s first historically.

However, when Mr Ganesan later says:

In height and the tip shapes, these signs should be made something larger than Sourashtra script Danda signs and different from Devanagari danda signs in code chart.

This we object to. There is no basis for such a distinct shape to the danda-s in Grantha. It irks us when Mr Ganesan makes such claims without providing proof. Where are the photos or textual references for the shape of the danda he prescribes? Without such proof, how is it that Mr Ganesan gives himself the authority to prescribe such a shape? On the other hand, these photos show that the same ordinary danda-s as are encoded in the Devanagari block are used in Grantha also:

1) from p 32 of Ārcika part of ref 3
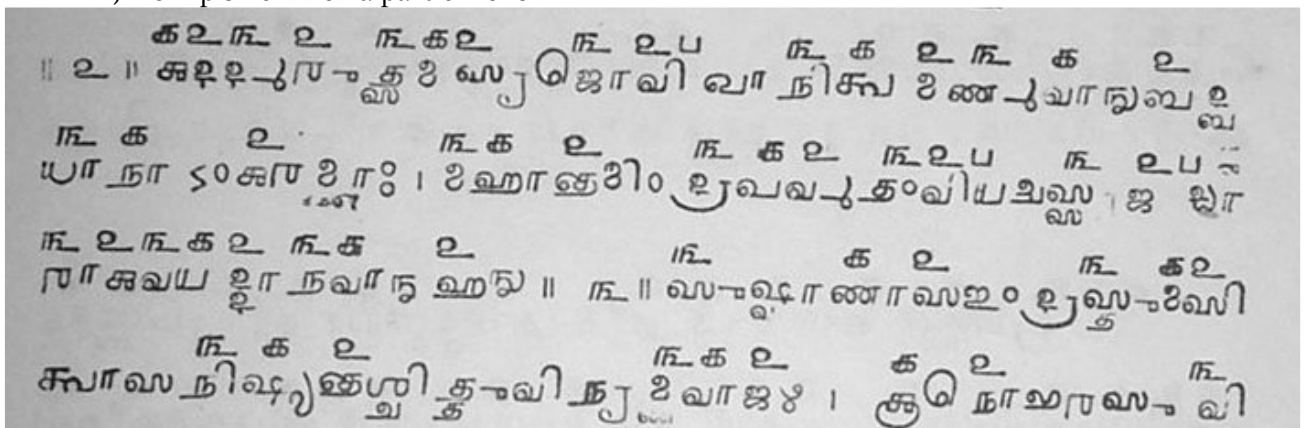
2) from p 1 of ref 7:
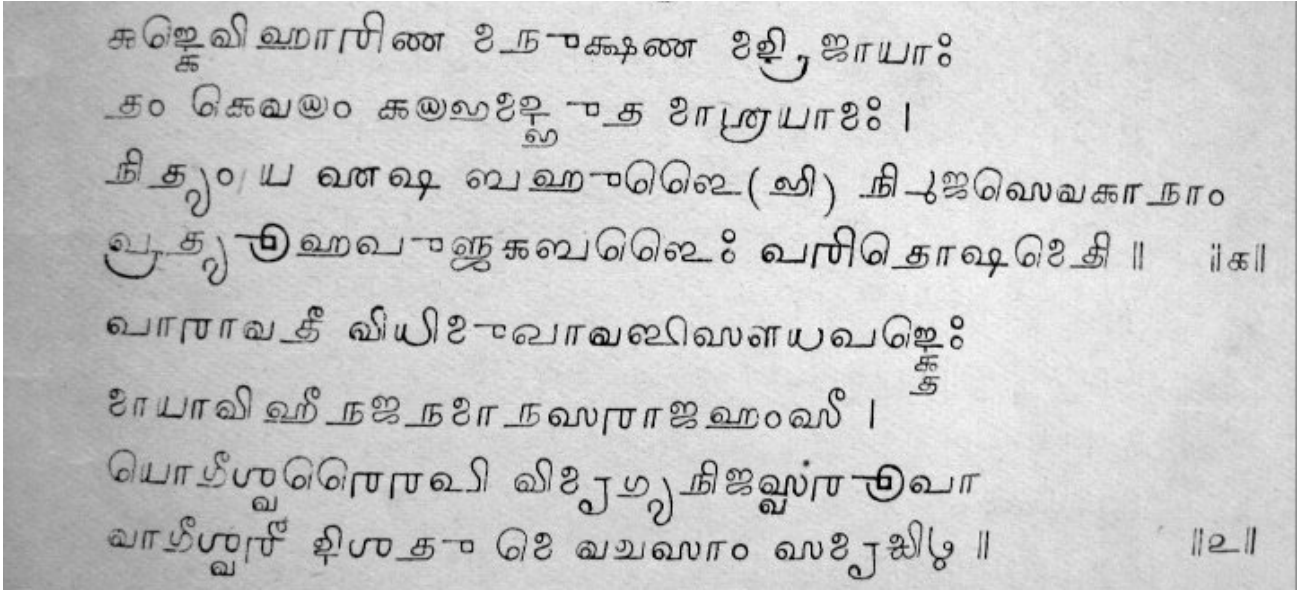
We are confident that Mr Ganesan can produce no other danda-s in Grantha which are distinct both in shape and function to which the following passage of the P&P document can apply:

> If and only if it can be demonstrated that orthographies using the existing script have a
> plain text contrastive use between two types of Dandas, use a combination of option a)
> and option b) above, to represent the distinction.

In summary, we object to encoding script-specific danda-s specifically because it will give rise to false claims such as Mr Ganesan's that the Grantha danda-s are graphically distinct from the danda-s used in all other Indic scripts. We therefore wish to follow the practice of all other non-Devanagari Indic blocks and reserve the positions corresponding to the danda-s with a recommendation to use 0964 and 0965 instead.

The P&P document which Mr Ganesan quotes uses the wording "generally taken as sufficient justification" which does not mandate the use of a script-specific danda in the case of encoding new scripts where danda-s are used. The same document continues:

> However, there may be considerations that would favor use of particular already-
> encoded Dandas from another block instead.

In the case of Grantha, the two considerations – 1. of asserting the identity of the Grantha danda-s with the Devanagari ones to prevent the emergence of any unauthentic script-specific danda-s (such as Mr Ganesan proposes) and 2. of consistency with the other Indic scripts – favour the use of the already-encoded danda-s from the Devanagari block. Further, the public review issue 59 on disunification of danda-s was closed four years ago (2005-05-18) in favour of unified danda-s, as said at http://www.unicode.org/review/resolved-pri.html. There is no reason to go against it.

Therefore script-specific danda-s should not be encoded. As in all other Indic scripts, the space should be reserved and a remark added to use the danda-s at 0964 and 0965.

III. OBJECTION TO ENCODING SEPARATE DIGITS AND NUMBERS

In support of encoding separate digits and numbers for Grantha, Mr Ganesan says:

> *Grantha digits are the original source numbers from which Tamil script numbers are a loan item. Grantha digits are more archaic in shape and functionality. Grantha numbers differ from Tamil numbers just as the Kannada and Telugu digits differ from each other. The shapes are different especially for 2, 3, 7 and 10.*

He claims that the shape of Grantha digits is 'more archaic'. First, we will ask the question: On what basis has Mr Ganesan distinguished 'Tamil digits' from 'Grantha digits'? He has not presented any proof to identify one set as belonging to one script and another set as belonging to another. We will however show that the exact same digits encoded in the Tamil Unicode chart are used for Grantha. Please see this photo from page 8 of the Table of Contents of ref 7 where we have marked the 'Tamil digits' 0-9 with circles, and where the place number system is also obvious.
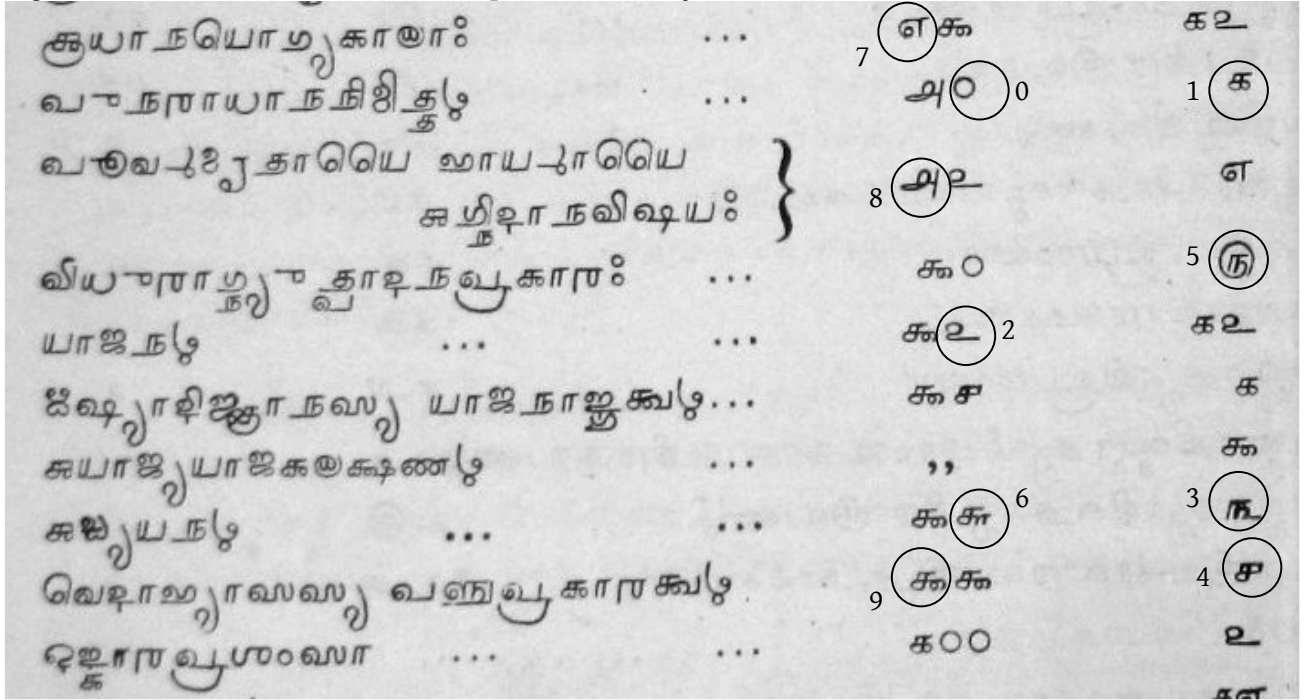
On the next page we also provide photos from pp 576, 577 of ref 8 from another publisher showing the same 'Tamil digits' being used. Any number of such photos can be given. It is therefore evident that the same glyphs shown as 'Tamil digits' are widely used in the context of the Grantha script too.

Of course, there are cases where the slightly different glyph for 3 that Mr Ganesan shows is used, as shown in the next page from p 100 of ref 9 (by a third publisher) with the 3 marked by arrows. There may also be cases where slightly different glyphs are used for the other digits (and numbers) too, such as the source Mr Ganesan has referred to as ref 6. However, these are merely glyphic variants, such as the Fraktur shapes of Latin letters, and hence do not merit disunification.

Even in Devanagari the contrast अ आ ओ औ ऋ ॠ ङ ण क्ष vs অ আ ও ঔ ঋ ৠ ভ ণ ন exists between two different styles of writing, but such different styles are not a sufficient reason for
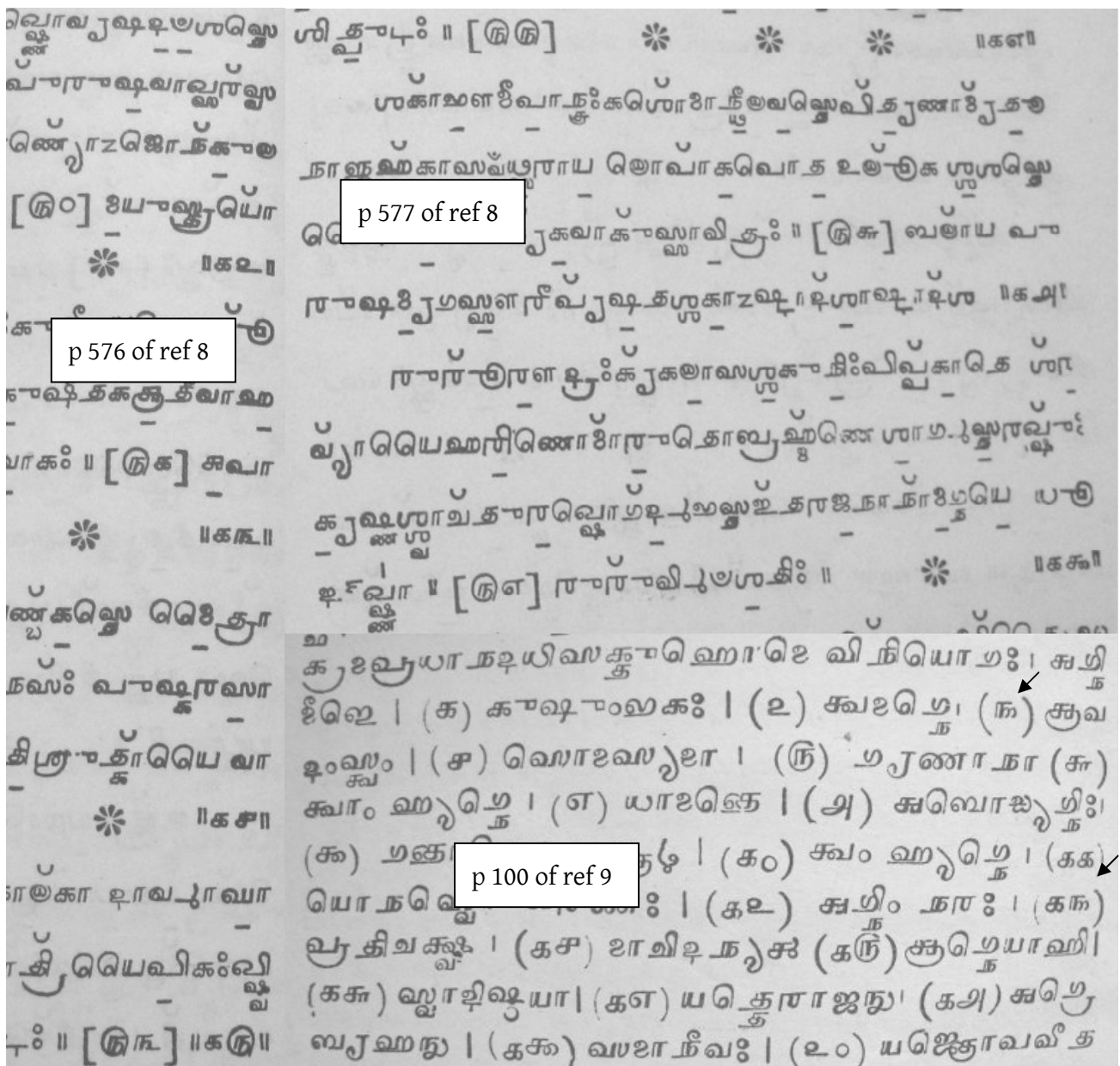
disunification. Such differences should be implemented at the font level, not the encoding level, as the question in the P&P §F.3.III prompts us to investigate: "Can the desired effect be achieved by changes to the display layer?" In fact, the the Devanagari contrast in the preceding text was achieved by using the two fonts Chandas and Uttara from ref 10. Similarly, the display of any 'archaic' forms of these 'Tamil digits' should be achieved by using a different font, not a separate encoding.

The claim that Grantha : Tamil :: Kannada : Telugu

Mr Ganesan says that Grantha digits and numbers are different from the Tamil ones in the same way that the Kannada and Telugu digits are mutually distinct. This is a false claim. Mr Ganesan has not shown proof that one set of digits belong to Grantha and another belong to Tamil. In the case of Kannada and Telugu, established practise uses a different *set of digits* for each script (despite some

digits looking the same in both) and any person acquainted with both scripts will easily identify one set as belonging to Kannada and the other belonging to Telugu. The same is not possible between Grantha and Tamil simply because the same set of 'Tamil digits' are widely used for Grantha, as we have shown. Therefore the Kannada-Telugu analogy does not work here.

And seeing as how there is widespread usage of the 'Tamil digits' for Grantha, one cannot disunify even though the usage context (i.e. the script of the text in which the digits are used) differs. A perfect example of this is how the Arabic-Indic digits are not encoded separately for Cyrillic, though they are used in Cyrillic texts too, simply because they are graphically and functionally non-different. (A search in the charts of the four Cyrillic blocks 0400–04FF, 0500–052F, 2DE0–2DFF and A640–A69F gives zero results for the word 'number' or 'digit'.)

This analogy with Cyrillic/Latin is particularly appropriate seeing as it also justifies the disunification for Grantha of the identical Tamil letters such as உ, உள etc (just as 0430 Cyrillic Small Letter A 'a' disunifies 0061 Latin Small Letter A 'a') while justifying the *non*-disunification of the Tamil digits. Cyrillic uses the same *set of digits* but not the same *set of letters* as Latin. The same is the case with Grantha and Tamil. Cyrillic encodes a separate small letter A since it would be awkward and pointless to have a gap in its set of letters. The same with Grantha. Cyrillic does not encode separate digits since the whole set of digits can be used as is from Latin. The same should be with Grantha.

While it can be argued that encoding separate numerals for Grantha is only a weak disunification as Grantha itself is new, it is a disunification nevertheless and so should not be done since it is not absolutely necessary. This principle is visible in the case of Cyrillic as shown above.

The case of the numbers 10, 100 and 1000

Mr Ganesan claims:

> *There is a variant shape, something like Greek omega letter, ௰ for Grantha Number Ten, However, in the proposed Grantha code chart, note the form taken from manuscripts and Grantha script books which is chosen to represent Grantha Number Ten glyph shape in standard Unicode Grantha font.*

Here Mr Ganesan has actually used the Malayalam number 10 0D70 in his PDF. He has not given any proof that it is a variant for Grantha number 10. We however, provide sample photos on the next page. The first one (on top) clearly shows that the 'Tamil' number 10 (circled and with arrow) has been shown with the exact same shape as in the Tamil block and after 8 and 9 (marked by arrows), so it cannot be suggested that this means something else in Grantha. The second photo shows the usage of 10 independently and in the old-style number 14 (10+4). These are also marked by arrows.

This proves that the ω-shaped glyph Mr Ganesan gives for the number 10 is just a glyphic variant, if at all it is attested, and hence not disunifiable. Mr Ganesan's number 100 also looks slightly different from the 'Tamil' number 100 but again, this is evidently just a glyphic variant. There are no claimed glyphic differences for the number 1000, and even if there are any, they are also variants.
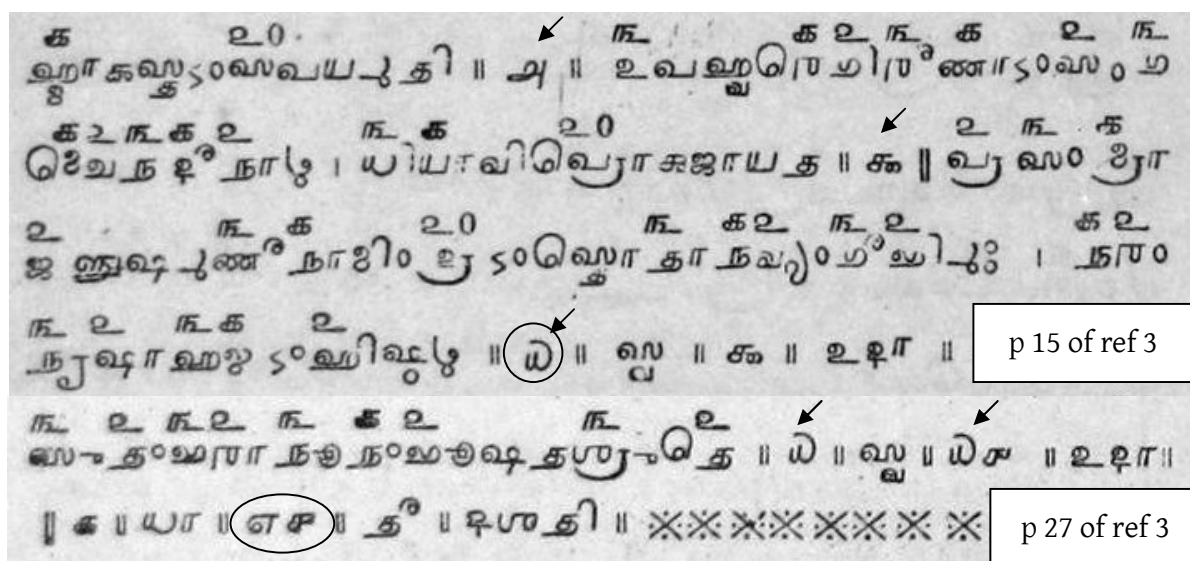
The claim of behavioural distinction of the 'Grantha digits'

Mr Ganesan claims that 'Grantha digits' are behaviourally distinct from 'Tamil digits':

The important point is that Grantha digits do not use the decimal place-value system at all where as Tamil digits use the place-value system. The Grantha system is based on a principle which is at once additive and multiplicative. To express multiples of tens, or hundreds or thousands in Grantha digits, the sign for 10, etc., is preceded by that of the corresponding units, which thus play the part of the multiplier. It is to be noted that Tamil digits do not employ this quasi-decimal system either in the web or in printed books, and hence encoding of Grantha digits is necessary for the multiplier basis of Grantha arithmetic.

We are indeed highly surprised that Mr Ganesan would write this in an official document submitted to the UTC. Does Mr Ganesan actually think that the UTC people would forget the TN#21 on Tamil Numbers: (http://www.unicode.org/notes/tn21/)? The change from the 'additive-multiplicative' system to the place-value system has occurred wherever the Tamil-Grantha digits are used, i.e. in the context of both Grantha and Tamil scripts, since the very many photos shown by us above clearly illustrate the place-value system being used in contemporary texts of Grantha. Even the source ref 3 which has used the 'Tamil' number 10 uses it only for numbers 10 to 19, as is visible in the second photo at the top of this page showing the number 74 (circled) written in the place-value system.

Further, Mr Ganesan's claim that 'Tamil digits do not employ this quasi-decimal system' is all the more surprising seeing as he himself has noted in his document http://nganesan.thamizamuthu.com/docs/Mal_0_numerics.pdf that Tamil and Malayalam formerly "employed numeric signs for 10, 100 and 1000 to write numbers larger than nine". If Mr Ganesan's words are intended to mean that *currently* Tamil uses the place-value system whereas Grantha uses the 'quasi-decimal' system, it is totally false. The photos we have shown are proof of that.

Thus there is no behavioural distinction of 'Grantha digits' and 'Tamil digits'. There is in fact no such thing as 'Grantha digit' separate from 'Tamil digit'. The same goes for the numbers 10, 100 and 1000. Therefore these digits and numbers must be properly appelled Grantha-Tamil digits and numbers, and their identity asserted by not encoding them separately in Unicode.

The IDN Security Question

Mr Ganesan says:

> *Also, for IDN and ability to protect from phishing, Grantha numerals can be separated*
> *from Tamil numerals by their atomic encoding in Unicode standard.*

We do not comprehend this sentence at all. How can 'atomic encoding' of Grantha numerals give protection from phishing? In fact, separately encoding what Mr Ganesan calls 'Grantha numerals' can only create a new security hole, since it adds to the number of ambiguous glyphs in Unicode.

In fact as we have said, there are no 'Grantha numerals' separate from 'Tamil numerals', and so we should not separately encode them in order to not increase the number of ambiguous glyphs in Unicode. (The ambiguity between Tamil and Grantha letters உ, ஊ etc cannot be helped, just as the ambiguity between 0061 Latin Small Letter A and 0430 Cyrillic Small Letter A etc cannot be helped.)

This being the case, what was Mr Ganesan thinking when he wrote that atomic encoding of Grantha numerals will protect from phishing? We suspect that since atomic encoding of chillaksharams was proposed as a solution for protection from phishing in Malayalam, Mr Ganesan has without due thought assumed that the same will be true for Grantha. We are surprised at this apparent carelessness of Mr Ganesan. The matter of encoding the Grantha script in the Unicode, which is going to be once and forever, must be considered very seriously and not carelessly.

Cost/Benefit Analysis

Despite there being no glyphic or behavioural difference, if another copy of the numerals from the Tamil block is placed in the Grantha block, the costs mentioned in the P&P §F.2, especially costs 1 (extra work for implementers) and 2 (confusion and mis-identification) occur. The latter cost also implies a security problem for IDN. On the other hand, there are no benefits to disunification.

Therefore the existing 'Tamil numerals' should not be disunified. The space they would occupy can be used for encoding other characters which are necessary and used in Grantha.

IV. OBJECTION TO ENCODING OM WITHOUT PROOF

Mr Ganesan proposes to encode a character which he calls the Grantha OM. We have seen the sequence OM represented in Grantha only as a O + MA + virama (usually rendered as a MA-chillu) or when followed by consonants, O + anusvara. We object to Mr Ganesan encoding a separate Grantha OM without providing attestation for the usage of the glyph he shows for that purpose. We do not have objection to the encoding of an OM for Grantha per se. However, so long as there is no proof for the usage of a particular single glyph for it, the codepoint should be left reserved but not encoded.

These photos from p 3 of ref 9 and p 1 of Gāna part of ref 3 show the current usage of OM:
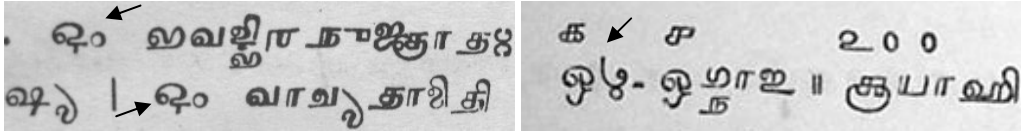
V. VOWEL SIGNS FOR SHORT E AND O

At the bottom of page 19 of his proposal, Mr Ganesan says that the short E and O vowel signs he proposes for Grantha are attested to by the work "Ancient and Modern Alphabets of the Popular Hindu Alphabets of the Southern Peninsula of India" of one Captain Harkness. He omitted to provide the URL http://www.archive.org/details/ancientmodernalp00harkrich from which this book is available for download, which would enable the readers of his proposal to learn that Captain Harkness has quoted these glyphs not for Grantha but for Tamil! The relevant parts from pages 1 and 2 of Harkness' work (note that the page numbering in the PDF seems faulty) are shown here:

Harkness has shown the Grantha vowels (both independent and dependent) exactly as they are written today. There are no glyphs shown for short E and O.

As for the independent vowels in Tamil: Short E looks like the current Tamil short E with the 'pulli'. Long E looks like the current Tamil long E. Short O looks like the Grantha O with the 'pulli'. Long O looks like the current Tamil long O.

As for the dependent vowel signs: Long E and O look like those used today for the short vowels. Those for short E and O correspond to the same but have the 'pulli'.

As such, this is a curious assortment that Harkness has presented for Tamil here! The only consistency we observe here is the usage of the 'pulli' to indicate short vowels. As such, though it is not a native Grantha usage, and though Harkness has not provided attestation for the usage of the Grantha independent vowel EE with 'pulli' even in Tamil, we observe that not much can be said against the 'pulli' being admitted as a short-vowel modifier for Grantha for the vowel signs E and O and the independent vowel EE, just as it has been used with the independent vowel OO.

VI. DRAVIDIAN CONSONANTS LLLA, RRA AND NNNA

Words and phrases from non-Sanskrit languages may require transliteration in Grantha, especially if Grantha and Sanskrit are to be promoted for modern usage. For this, Mr Ganesan has

suggested importing the old forms of the Tamil vowels short E and O and the letters LLLA, RRA and NNNA. We have discussed the vowels. Now we consider the consonants.

Here we do not believe that Tamil needs to be transliterated into Grantha, despite Mr Ganesan's alleged samples from one Samskrita Granthalipi Sabha, Chennai. Mr Ganesan did not give us, despite requests, contact information for said Sabha and our own inquiries were not fruitful. Therefore we doubt the authenticity of these samples, especially since they are in handwriting.

There are few people, if at all any, who can read Grantha but not Tamil. Anyone who can read Grantha can easily learn to read Tamil, just as is the case in Kannada and Telugu. Wherefore the need to write Tamil in Grantha? This is quite unlike the situation where there can be a valid need to write Tamil in Devanagari since most people who can read Devanagari cannot read Tamil. Writing Tamil words in Grantha can only be for amusement and so does not warrant encoding.

Therefore the only valid case for inclusion of the Dravidian consonants (and the short vowels) is when words or phrases from Malayalam, Kannada or Telugu words need to be written in a Sanskrit text written in Grantha. Any characters required for this purpose can be used as-is from the Tamil block. Mixed usage of Tamil and Grantha scripts is well known in the case of writing Manipravalam where Tamil and Grantha characters are mixedly used even within a single word but retaining their native appearance and behaviour. Tamil characters should therefore not take up Grantha vowel signs or Grantha-style consonant cluster behaviour, since that is not attested in Grantha. The corresponding positions in the Grantha block may be reserved.

Thus these characters should not be disunified since mixed Tamil-Grantha usage already exists, and the required glyphs are the same as in Tamil, unlike the different glyphs for short E and O.

VII. REFERENCES

1. Taittirīya Brāhmaṇam, Heritage India Education Trust, Chennai, 1980.
2. http://sanskritweb.net/yajurveda/ta-03.pdf, retrieved 2009-Aug-14.
3. Sama Veda Samhita, AVMSSGDPS Samithi, Kumbakonam, Tamil Nadu, 1985.
4. http://is1.mum.edu/vedicreserve/sama_veda/sama_veda.pdf, retrieved 2009-Aug-14.
5. South Indian Scripts in Sanskrit Manuscripts and Prints, Reinhold Grünendahl, Wiesbaden, Germany, 2001. ISBN 3-447-04504-3.
6. http://unicode.org/review/pr-37.pdf, retrieved 2009-Aug-14.
7. Smṛti Muktāphala – Varṇāśrama Dharma Kāṇḍa, Sharada Vilasa Press, Kumbakonam, Tamil Nadu, 1939.
8. Taittirīya Saṃhitā, Heritage India Education Trust, Chennai, 1980.
9. Āśvalāyana Śrāddha Prayoga, AVMSSGDPS Samithi, Kumbakonam, Tamil Nadu, 1985.
10. http://sanskritweb.net/cakram/index.html, retrieved 2009-Aug-16.
11. Grantha Bala Pathamala, Heritage India Education Trust, Chennai, India, 2003.

–o-o-o–