Date: Sun, 02 Aug 2009 09:55:12 +0530
From: Shriramana Sharma <jamadagni@gmail.com>
To: Rick McGowan <rick@unicode.org>
CC: "N. Ganesan" <naa.ganesan@gmail.com>,
        Elmar Kniprath <kniprath@online.de>
Subject: Grantha proposal

<div style="border:1px solid black; display:inline-block">**L2/0-9-258**</div>

Hello. Thanks for your reply. I have maintained the CC to Magda Danish. Please remove if
not necessary. OTOH please maintain CC to my friend Elmar Kniprath with whom I have been
working on the past few months for a Unicode encoding for Grantha independent of and
ignorant of the work of N Ganesan.

> If you have comments for UTC consideration, please feel free to submit
> them to me, and I will see that they are put on the agenda for the next
> possible UTC meeting. The document deadline for the upcoming UTC meeting
> is August 3, but I can receive and post "comment" documents until August 7.

Thank you for your offer. My comments on Mr Ganesan's proposal are as follows.

I. On 1134E Grantha Chillu Marker

I understand that Mr Ganesan advocates the addition of this separate character which he
argues, is necessary for the formation of the various chillaksharams that were used in
Grantha script. However, I advocate that this is NOT necessary. My arguments are as
follows:

Mr Ganesan gives as supporting evidence for the inclusion of a separate chillu marker the
fact that some Malayalam chillaksharams have received separate codepoints in Unicode
5.1.0. I have read the relevant document http://www.dkuug.dk/jtc1/sc2/wg2/docs/n3126.pdf
and I find that a similar case can NOT be drawn for Grantha.

The main argument that is present in the above document for the separate encoding of the
Malayalam chillaksharams is that the chillu form of some consonants is NOT semantically
equivalent to the corresponding "killed-vowel" or "dead" forms of the consonants formed by
adding a virama. So:

(van-yavanika == big curtain) != (vanyavanika == wild forest)

However, there is no such demonstrable contrast between the chillaksharam form of
consonants and the "dead" form of consonants in Grantha. I encourage Mr Ganesan to provide
such proof if he has any. In the meanwhile, I speak as a person who daily uses Grantha to
read Vedic texts and to write when I teach Sanskrit/Vedic lessons to my students that
there is absolutely no ***semantic*** contrast between consonant + virama and
consonant-chillu.

The only place which may be said to specifically require the chillu form of a consonant to
be displayed is in the case of the Taittiriya School of the Krishna Yajur Veda, for whose
books the Grantha script is still widely used by teachers and students today. Even that is
only in the case of the consonants velar nasal NG, dental nasal N, geminated dental nasal
N + N, and the trill R, and that too only when there is a certain ***phonetic*** quality of
those sounds to be denoted, not because there is any ***semantic*** difference.

Please see the attached image. As far as the Vedic/Sanskrit ***language*** is concerned, each
of the various forms presented under 1, 2, 3 and 4 is ***semantically*** the same. There is NO
difference in meaning, UNLIKE van-yavanika / vanyavanika in Malayalam.

The phonetic difference, however, can be summarised thus: when the "dead" consonant C1 is
displayed separately, i.e. without forming a conjunct or sub-base/post-base conjoining
form with the following consonant C2, the pronunciation of C1 should be followed by a
small gap before the pronunciation of C2. That is, to indicate this small gap in
pronunciation, C1 is written/printed separately.

Thus 1a and 1b both denote the pronunciation with a gap between NG and V whereas 1c
denotes one without a gap. Next is 2a against 2b/2c. Next is 3a against 3b and then 4a
against 4b.

Note that in 1a the chillu form of NG has been used and in 1b the NG + VIRAMA form has
been used (older printings used the chillu and newer printings the virama form) with both
the same semantic ***and*** phonetic value. The point is that NG stands independently without
combining with the following VA, and whether this is achieved by a chillu or by a virama
is immaterial.

Therefore, there is no semantic contrast between the chillu form or virama form of a
consonant. And even without a separate chillu marker, one can achieve the rendering effect
of an independent chillu without conjoining as follows:

1. Font makers should provide all useful chillu forms and rendering engines should display
the chillu form of C1 in the place of C1 + VIRAMA where such form exists. I say "useful"
because of all the chillus presented by Grunendahl, only NGA, TTA, TA, NA, MA (and the
chillu of the consonant cluster NA + VIRAMA + NA) have practical usage today, and can be
recognized by current users of Grantha.

2. By the rules outlined at http://unicode.org/review/pr-37.pdf page 14, C1 + VIRAMA + C2
would cause the full form of C1 to be used with the sub-base conjoining form of C2 being
attached to it. In cases where we wish the chillu should be used in such a sequence, we
should insert a ZWNJ so that: C1 + VIRAMA + ZWNJ + C2 ensures that the VIRAMA does not
cause C2 to be displayed in the sub-base conjoining form but the C1 + VIRAMA sequence is
rendered as the chillu.

When there are two different ways of writing the same content, for example, see
http://sanskritweb.net/cakram/index.html which shows in the pictures below chandas.ttf and
uttara.ttf the two different styles prevalent in writing "a" "aa" etc, the only proper
thing to do is to resort to two different fonts. So if Mr Ganesan envisages a situation
where he wants to show the purely orthographic distinction between using the overt virama
form and the chillu form, rather than asking for a separate place in the Unicode character
space for the two different forms, he should use two different fonts one of which has the
chillu form and the other does not.

Thus I do not see any need for a separate chillu marker. Introducing one would even create
the misconception that there exists some semantic difference between the chillu and the
virama and hence should not be done.

II. On 11302 ANUSVARA and 11303 VISARGA

It is better if the anusvara and visarga are encoded with general category Lo rather than
Mc (as hinted by the dotted circle before the visarga in Mr Ganesan's proposal) because
they denote independent phonemes, and ignorance on the part of rendering engine developers
as to to which characters these can potentially attach themselves results in situations
where one is not able to form a proper rendering.

See the attachments independent-anusvara.jpg and independent-visarga.jpg of the desired
appearance of anusvara and visarga which would be difficult to achieve if the rendering
engine people did not do their work correctly.

These are just two of many passages in the Sama Veda where the anusvara and visarga are
written after one or more numbers. One could only guarantee the absence of any problems in
such situations if the Unicode people themselves label the symbol as Lo instead of Mc.

(I have asked for the category of the Devanagari visarga to be changed, too, but have not
done the same for Devanagari anusvara because it is by nature not a spacing character and
one would be hard put to place it after a number in a similar context. A Devanagari book
of the Sama Veda actually uses the number zero after the number 3 to achieve this effect!)

III. On 11362 VOWEL SIGN VOCALIC L and 11363 VOWEL SIGN VOCALIC LL

The dotted circle for these is missing in the current proposal draft so please ensure that
they carry the property Mc with combining class 0.

```
IV. On 11364 DANDA and 11365 DOUBLE DANDA
------------------------------------------
```

I thought that the general policy was that while the Devanagari danda-s exist, separate
danda-s are not going to be allowed for other scripts in the absence of a significant
difference. It is obvious from the long printed passage whose picture is included in page
16 of Mr Ganesan's proposal that there is no special shape to the Grantha danda-s compared
to the Devanagari danda-s. Therefore these two should not be separately encoded.

```
V. On the digits 11366 to 1136F
-------------------------------
```

These are the same as the "Tamil" numbers (which should properly be called the Grantha
numbers) encoded at 0BE6 to 0BEF. So one can only say that a good Grantha font should have
those "Tamil" numbers, and one should not encode these separately for Grantha.

```
VI. On the numbers 11370 to 11372
---------------------------------
```

These are also identical with the Tamil numbers 0BF0 to 0BF2, and the slight difference in
appearance is no doubt due to the various stylistic difference seen in the ancient Grantha
inscriptions or perhaps merely the script itself evolving. In any case, they do not have
any semantic difference nor any significant orthographic difference from those Tamil
numbers and hence should not be encoded separately.

```
VII. On the DIRGHA SVARITA svara mark 11355
-------------------------------------------
```

A note must be added to the effect that this marks all kinds of svarita (i.e. both
"normal" svarita and dirgha svarita) in Rig Vedic texts.

```
VIII. On the signs 1135A and 1135B
----------------------------------
```

There are four different types of anusvara-s in the Krishna Yajur Veda.
These are detailed at my paper on the subject at
http://sanskritweb.net/sansdocs/index.html#IPA. I will merely call them anusvara 1-4 here.
Anusvara-1 (may be approximated as "um") is denoted in Grantha by 11302. Anusvara-4
("ggu") is denoted by 1135B. Both anusvara-2 ("gum") and anusvara-3 ("gu") are denoted by
1135A.

See the attachment anusvara-type-2.jpg. In line 1 and 2, 1135A indicates anusvara-2 "gum"
which is pronounced with a nasalization at the end. On line 5, 1135B indicates anusvara-3
"gu" which does not have a nasalization at the end. Whether anusvara-2 or anusvara-3 is
meant by the symbol can easily be determined by context, as detailed in my paper. When a
conjunct consonant follows, anusvara-3 is to be pronounced. Otherwise anusvara-2 is to be
pronounced.

Now the point for our discussion is that the labeling of 1135A as GRANTHA SIGN GUM is not
very satisfactory, since when it denotes anusvara-3, the pronounciation is without the
nasal. There even exist some minor traditions where anusvara-2 as outlined in my paper
does not exist at all and is totally replaced by anusvara-3. Therefore such traditions
would never pronounce the nasal and the labeling of "GUM" is totally misleading.

So I suggest that these two signs merely be labeled as:

```
GRANTHA SIGN VEDIC ANUSVARA 1
GRANTHA SIGN VEDIC ANUSVARA 2
```

```
IX. On Vedic Svara symbols
--------------------------
```

Mr Ganesan has shown only the three Vedic Svara symbols in his proposal - ANUDATTA,
SVARITA and DIRGHA SVARITA. These are sufficient for the presentation of the Rig Vedic and
Yajur Vedic texts in Grantha. (No known Atharva Vedic text in Grantha exists, though that
could be done too using the same three svara symbols.)

The Sama Veda, however, is a different world. I, myself a scholar of the Yajur Veda, have done research into the svara symbols used in Grantha writing/printing of the Sama Veda with the help of a relative of mine who is one of the foremost scholars of Sama Veda in South India.

I attach the file sama-veda-svaras.pdf which is an extract from a draft shared by Mr Kniprath with me. Please see. Some of these symbols overlap with those detailed by and proposed by N3366 (later slightly amended by N3383). In those cases, they had better be encoded in the Vedic Extensions block at the end of 1Cxx, while ensuring that they can be used with Grantha symbols also.

X. On inclusion into the BMP

In this entire document, we have alluded to the Grantha characters by the codepoints in the SMP range 113xx because that is what Mr Ganesan has done and what is tentatively allocated to Grantha in the page http://www.unicode.org/roadmaps/smp/. However we would also prefer inclusion into the BMP as Grantha is still a living script in widespread use in many parts of Tamil Nadu (and even some scholars in Karnataka and Andhra use it).

Also currently many Unicode-aware systems support only 16-bit Unicode, i.e. use 16 bits to represent individual Unicode codepoints, making them incompatible with the SMP etc. Of course that is not something for the Unicode people to worry about, but still it would speeden up the adoption of Grantha if it would not be put into the SMP but put into the BMP.

However, seeing as how most of the BMP has been already hogged by the CJK ideographs and the Hangul syllables, with the remaining space being taken up by the already encoded scripts I don't see that a fully unoccupied 128-codepoint range can be allotted for Grantha in the BMP. :(

There are only two possibilities for the Grantha script to be encoded into the BMP: one is to take away some space from the PUA E0xx to F8xx but I am not sure that the Unicode people will be willing to do that.

The other chance is: if the space 1C80 until before the Vedic Extensions block (as shown in http://www.unicode.org/roadmaps/bmp/) is allocated, by pushing the Sund script which seems to require only a small codespace to some other space like at the end of A9xx, we could encode Grantha easily in a contiguous block of the BMP.
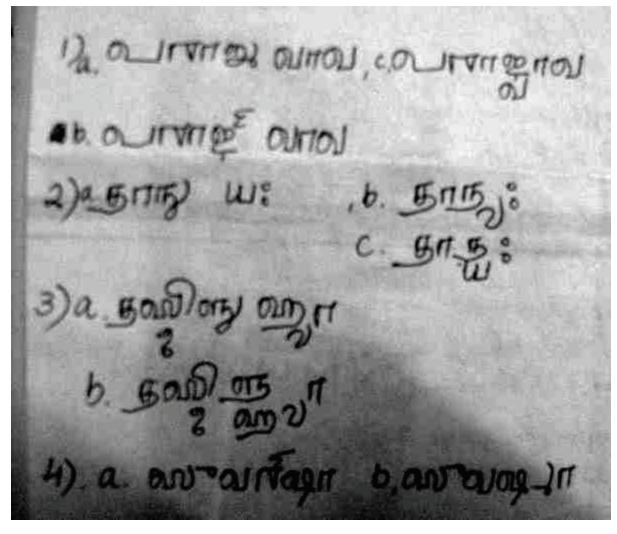
Some of the Sama Vedic signs which are not script-specific (in appearance) can be accommodated in the Vedic Extensions block. Others which are specific to Grantha (like the superscript numbers 0 1 2 etc) can be included in the Devanagari Extensions block by renaming the name of that block to Vedic Extensions 2 (Script Specific) or something like that, since these symbols share the same semantic sense as those Devanagari-specific signs.
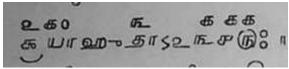
By this the Grantha script can be included fully in the BMP without hindering anyone else. Just a small courtesy of accomodation is required from those who have proposed the inclusion of the Sund script in that their script needs to be moved from later part of 1Cxx to end of A9xx.
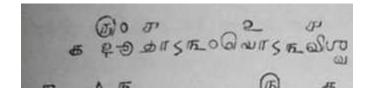
XI. Conclusion

Given a couple of weeks time, during which I am obliged by my employer to finish off another high-priority work, I will provide a polished proposal which includes all the good work done by Mr Ganesan, and the research and work done by Mr Kniprath and me. In the meantime I request the Unicode people not to hurriedly pass the current proposal as the final ultimate Grantha specification.

Shriramana Sharma.

1) a. பொனாஜு வாவ, c. பொனாஜாவ வு

b. பொனாஜ் வாவ

2) a. தாரு ய: b. தாரு:

c. தாது:

3) a. தவிினு ஹூா

b. தவிினு ஹுா

4) a. லௌவஸ்ஷா   b. லௌவழி_ா

உ கெ0   கூ   க கக
கூ யா ஹூ-் தா ஜெ உ ரூ சு ரூ: ா

ஜ்0 சு   உ சு
க உஉ கா ஜ கூ0 வொ ஜ கூ விஎ வ

நாநிவாஷிஷ்ணயாசு । மாயத்யமஹொ

தஸ்யொஷ்ஸம்வாாவததசு । [சங] தகூ

நு । உநெராவபதுஹளு । தஸ்வஸ்வராாவ

ஹா நாநுஇவஷு । வமுவொவெஹா

வெஷாோாஜா । யஊாஊாசாஸ்ஸு ஹா நா

ணாதி । ஹாஊௌவ ராஜாநஉஷி ஷ்ணா

தநபூாதஸ்வெநஸ்ரீணீயாசு । உஊாஊ ஞ்நெ ।

தஷிபெரணதுதீயஸ்வவெ । சஷிவெஷாஊவெஷ்ஸ

தெராஹாஊ । யஹனவ௷டுஜொவபதாஸ்யம் । த

யெயாம் । னஸ்காாாணஷிணாஊவாவெத்�௷வெணவ ।

- sign invoking the god Ganesha

xx60 ◌ GRANTHA ACCENT SIGN ANUDATTA

- may also be used with non-superscript digits

xx61 ◌ GRANTHA ACCENT SIGN SVARITA

- may also be used with non-superscript digits

xx62 ◌ GRANTHA ACCENT SIGN DIRGHA SVARITA

- in Rig Vedic texts also used for SVARITA instead of xx61
- may also be used with non-superscript digits

xx63   \<reserved\>

xx64   \<reserved\>

xx65   \<reserved\>

Digits

As Grantha digits are identical with Tamil digits a Grantha font must contain the Tamil digits.

Various signs

The following signs may occur as singles or combine with eachother in various ways to denote Vedic accents and tones. They may also be used with non-superscript digits.

xx66 ◌ GRANTHA SUPERSCRIPT ONE

- Sama Veda accent sign Prathama
- Sama Veda accent sign Udatta
- also used sometimes to indicate vowel length

xx67 ◌ GRANTHA SUPERSCRIPT TWO

- Sama Veda accent sign Dvitiya
- Sama Veda accent sign Udatta
- Sama Veda accent sign Svarita

xx68 ◌ GRANTHA SUPERSCRIPT THREE

- Sama Veda accent sign Tritiya
- Sama Veda accent sign Anudatta

xx69 ◌ GRANTHA SUPERSCRIPT FOUR

- Sama Veda accent sign Chaturtha

xx6A ◌ GRANTHA SUPERSCRIPT FIVE

- Sama Veda accent sign Mandra

xx6B ◌ GRANTHA SUPERSCRIPT SIX

- Sama Veda accent sign Atisvarya

xx6C		<reserved>
xx6D		<reserved>
xx6E		<reserved>
xx6F	◌	GRANTHA SUPERSCRIPT ZERO

- Sama Veda sign non-prolated
- marker for some special kinds of Svarita

xx70 ◌ GRANTHA SUPERSCRIPT TAMIL PA

- Sama Veda accent sign Dipta Visvara
- Sama Veda accent sign Shivaka Udatta

xx71 ◌ GRANTHA SUPERSCRIPT KA

- Sama Veda accent sign Kampa
- marker for independent Svarita

xx72 ◌ GRANTHA SUPERSCRIPT NA

- Sama Veda tone sign Namana Visvara

xx73 ◌ GRANTHA SUPERSCRIPT A

- Sama Veda tone sign Abhigita Visvara

xx74 ◌ GRANTHA SUPERSCRIPT VI

- Sama Veda tone sign Vinata Visvara

xx75 ◌ GRANTHA SUPERSCRIPT STAR

- Sama Veda tone sign to denote Independent Svarita in modern printings

xx76 ◌ GRANTHA SUPERSCRIPT DASH

- Sama Veda tone sign Prenkhana Visvara

xx77 ◌ GRANTHA SUPERSCRIPT AYATA

- Sama Veda tone sign Ayata Visvara

xx78 GRANTHA SUPERSCRIPT NONSPACING DOUBLE BAR

- Sama Veda Section Separator, non-stopping
- to be inserted between two writing syllables

Subject: Re: [Fwd: Subj: Unable to download Grantha proposal]
From: Shriramana Sharma <jamadagni@gmail.com>
Date: Sun, 02 Aug 2009 15:54:00 +0530
To: Rick McGowan <rick@unicode.org>
CC: "N. Ganesan" <naa.ganesan@gmail.com>, Magda Danish <magda@unicode.org>, Elmar Kniprath <kniprath@online.de>

```
Hello. I just wanted to add what I forgot to mention in my last mail:

1134C VOWEL SIGN AU is shown as a two-part sign. However, in current Grantha usage only
the right-hand-part of these two parts is used. The situation is perhaps analogous to that
of the Malayalam 0D4C.

Shriramana Sharma.

.
```