Dear Michael,

Below I ask questions regarding the display of alternate glyphs important for those who work with Sanskrit and Vedic. Please clearly explain the technical points for me so that I may clarify them for non-technically-savvy users. I also propose alterations to our Vedic Unicode Proposal draft.

I.   Glyph variants of characters

I understand how a font will allow both variants of each of the following characters to be used in the same document:

a, *ā*, *r̥*, *r̥̄*, *jh*, *ṇ*, and *ए*?   Variants of these characters need not be encoded as independent characters in the Devanāgarī code block.

Any editor that uses Apple Advanced Typography API (Application Programming Interface) allows variant shape selection. On Mac OS X, Text Edit has under the Format menu under the Font submenu, a section containing the commands Baseline, Kern, Ligature, and Character Shape. The Ligature command allows one to select alternate ligatures such as vertical versus horizontal conjuncts, and an unconjuncted sequence of characters with virāma between. The Character Shape command allows one to select alternate character shapes available in the font such as varieties of *n*.

Alternate shapes such as the long s were made separate characters only if they had distinct significance in some context or other. Long s is used in IPA for an alveolar s. Likewise Turkish uses both dotted and undotted i, each representing a different phoneme.

II.   YV svarita characters

The situation of two of the YV svarita characters is just like the situation of alternate character shapes in point I. We should eliminate one of the two characters 1A38 and 1A3A in Table XX: Vedic Extensions. A Vedic font that includes both the elbow glyph and the umbrella handle glyph should allow character shape selection of one or the other. In an email to you dated 20 January, I argued the following in my second point:

We found the umbrella handle below (1A3A) in the *ḫājasaneyisaṁhitā* in identical contexts with the elbow (1A38) confirming your initial hunch that it was a graphic variant (and not, as I argued, a sign specific to the Kāṭhaka Saṁhitā). There should be just one sign either character 1A38 or 1A3A. However, the umbrella handle seems to be more primary and the elbow seems to be a typeface imitation of it. Lets keep the umbrella handle image and eliminate the elbow image.

Both glyphs should be included in an adequate Vedic font. Opposite to what I wrote, the elbow should probably be the standard image since it is used in printed texts. Please eliminate 1A3A in Table XX: Vedic Extensions and adjust the table accordingly.

III. Kṛṣṇa Yajur Vedic Maitrāyaṇī and Ṛgvedic Śāṅkhāyana dependent mid-tone svarita

The horizontal stroke through a character at mid-height for the Maitrāyaṇī/Śāṅkhāyana dependent svarita ought to get positioned properly, which requires different height adjustments for different characters, should be independently selectable for different font coloring, and should be a combining character to prevent formatting interruption.

The horizontal stroke should pass through a character at a height where it does not interfere with diacritics and is not confusable with other glyphs. For example, through a *t* त it should pass through the vertical bar alone below the head bar but above the distinguishing mark. In this way it will not be confused with double *t* त्. Likewise the horizontal accent mark should not pass through the diacritic dot in velar *n* ङ but above it.

Moreover, such characters cannot be made simply with the strike-through font option because they should be independently selectable so that they can be colored red. Accent marks are typically shown in red in manuscripts.

There already exist two strike-through characters in Unicode: Stroke Overlay Combining Short (non-connecting) [0335], and Stroke Overlay Combining Long (connecting) [0336]. Font designers should use the short one for the mid-height horizontal stroke Maitrāyaṇī/Śāṅkhāyana dependent svarita tone. They should create many strike through glyphs in the font and write font ligation rules to place them correctly. The font should NOT make precombined characters. Manuscripts typically mark accents in red, while the accented text appears in black. The accent mark must remain a separate character so that it will be possible to select and change the color of the strike-through character to red without changing the color of the akṣara itself.

The horizontal accent stroke is not a strike through font formatting option, but it need not be a special character proper to Vedic encoded in the Vedic accent section with other svaritas, [in particular, one does not need a separate character Maitrāyaṇī mid-tone svarita right after 1A33, which is a Maitrāyaṇī triple-tone svarita].

IV. Triple svarita

Please modify the images of the svarita [1A71] in the Devanāgarī Table and the double svarita [1A32] and triple svarita [1A33] in the Vedic Extensions Table so that the vertical lines are rectangular, not wedges.

V. Nasals

A. *Phonetics*

1. The following distinctions in nasals are accounted for in ancient Indian phonetic treatises (Prātiśākhyas):

   a.   nasalization of vowels

b. nasalization of semivowels *y v l*.

c. short anusvāra, which occurs after long vowels

d. long anusvāra, which occurs after short vowels

According to medieval phonetic texts (Śikṣās), the following additional distinction is made:

e. heavy anusvāra, which occurs before a conjunct consonant

f. two-mora anusvāra.

2. The following distinction in nasals is accounted for in South Indian Taittirīya Yajurvedic recitation:

Three types of *anusvāra*:

a. g before spirant initial in a conjunct, after a long vowel.

b. gg before spirant initial in a conjunct, after a short vowel.

c. gM before *r*, spirant or before a vowel.

In Tamil script, the distinctions in A2 are shown by using regular characters, as they are in Kannada and Telugu scripts (see TSTBNotesV.rtf, section VIII), and transcriptions of these texts into Devanāgarī do the same. As shown in the document VedicMarks2007Feb23V.rtf, the *g* ग्, double *gg* ग्ग्, and *gṁ* ग्ं are used for A2a, b, and c respectively.

No further phonetic distinctions are prescribed by the *Nāradaśikṣā* 2.4.5-7, except by 2.4.6cd which prescribes shortening of an *n* by a quarter mora and nasalization of the preceding vowel by half a mora (see Vedic Marks2007Feb23V.rtf II.3. This however has not been found represented in any texts.

B. We find the following marks representing nasals in Vedic texts:

1. *Vājasaneyisaṁhitā* (see the document PassagesVajasaneyiEdited.rtf, sections labeled "Nasalized Vowel," "Nasalized Semivowel," and "Anusvāra")

a. Bindu over the preceding vowel where the *Vājasaneyiprātiśākhya* prescribes nasalization of the following semivowel *y v l* (A1b). (See section "Nasalized Semivowel" and document SBrNotesV.rtf, paragraph beginning, "According to VPr. 4.10")

b. Candrabindu over a vowel where the *Vājasaneyiprātiśākhya* prescribes nasalization of the vowel. (See section "Nasalized Vowel" and document SBrNotesV.rtf, paragraph beginning, "There are instances in"). (Similarly, see TSTBNotesV.rtf "Candra-bindu is used over the vowel preceding double l".)

c. Hanging *n* or six dot where the *Vājasaneyiprātiśākhya* prescribes long anusvāra.

d. Antargomukha where the *Vājasaneyiprātiśākhya* prescribes short anusvāra.

2. *Śatapathabrāhmaṇa* (see the document SBrNotesV.rtf).

a. Gomukha with bindu, gomukha with candrabindu, and gomukha with candrabindu and virāma all representing anusvāra without distinction.

b.   Six dot is used here also without distinction. [The usage here does not correspond with that in the *Vājasaneyisaṁhitā* where it occurs only after a short vowel to indicate a long anusvāra; in the example shown it occurs after a long vowel.]

c.   Candrabindu with the numeral 2 or 3 below after a vowel and followed by a double danda and avagraha, where VPr. prescribes a nasalized vowel. (See the document SBrNotesV.rtf, paragraph beginning "Candra-bindu with the number two or three below".)

3. Taittirīyā Śākhā (See document TSTBNotesV.rtf)

a.   Bindu used to represent a nasalized stop or semivowel homorganic with the following stop or semivowel. (Shown in (1).)

b.   Candra-bindu with virāma is used for anusvāra or nasalization before *r*, spirant. (Shown in (2).) This corresponds with the phonetics described above in A2c. Witzel mss. cited in TSTBNotesV.rtf V and VI use this sign for the next as well without distinction, as does Penn ms. 2032 (See TSTBNotesV.rtf VII. See also Whitney's comment in IX.)

c.   Candra-bindu over candra-bindu with virāma is used before a spirant initial in a conjunct consonant. (Shown in (3).) This corresponds with the phonetics described above in A2a or b.

d.   Candra-bindu above an akṣara followed by the numeral 4 denoting the dīrgha raṅga kampa, i.e. prolonged nasalized vowel, which is recited with the tone of an aggravated svarita.

4. Ṛgveda

a.   Candrabindu with avagraha below. (See RgvedaUPenn.rtf UPenn No. 15.)

b.   Candrabindu with numeral 3 below. (See RgvedaUPenn.rtf UPenn No. 2021). Sāyaṇa unambiguously states that this is a prolonged nasalized vowel.

C. Conclusions regarding Nasals

1. Six dot is synonymous with rthang

Vedic Extensions 1A49 and 1A4A are graphic variants of candrabindu with hanging *n*. This is demonstrated in MS Indic 371 folio 3 recto where variants of a cross between candrabindu with hanging *n* (row 1) and rthang (row 2) before *sarvva* (generally before spirants and *r*) are found. The hanging *n* is synonymous with the six dot grapheme in representing a long anusvāra, as is shown clearly and succinctly in the section "Long (preceded by a short vowel)" in the document "PassagesVajasaneyiEdited.rtf". Hence in the Vedic Extensions Table, eliminate 1A4A rthang, which is a graphic variant of 1A49 six dot, and replace the image of 1A49 by one that looks more like the chandrabindu *n* in MS Indic 371 folio 3 recto row 1, or in DRGG27.40 line 1.
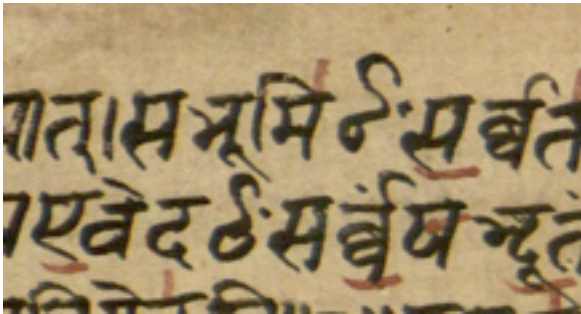
2. The various gomukhas are merely graphic variants of each other.

The various gomukhas slated to be encoded in the Vedic Extensions Table 1A46, 47, 48, have no distinct significance; they are merely graphic variants of each other. It seems then that only one ought to be encoded and fonts should include the varieties as shape variants.

3. Candrabindu combines with various marks below.

Candrabindu combines with various marks below some of which are slated to be encoded in the Vedic Extensions Table:

a.   virāma (1A43)

b.   candrabindu virāma (1A44)

c.   the numeral 2 (1A45)

d.   the numeral 3 (missing)

e.   avagraha (missing)

Although a, b, and c are found with multiple significance (candrabindu with 2 below is used synonymously in some ms. with double candrabindu virāma; both are actually synonymous with the six dot and rthang for long anusvāra), in identical contexts with c, d, and e, in other texts are seen candrabindu over the preceding syllable with the numerals 1, 2, 3, 4, or avagraha following the syllable. This in itself hints that the candrabindu indicates one thing, and the numeral or avagraha another. Sāyaṇa's comment on B4B affirms this conclusion definitely. The candrabindu indicates nasalization of the preceding vowel just as it does if it occurs above the vowel, and the numeral indicates the length of the preceding vowel, just as the numeral 3 does in the ordinary indication of prolongation. Therefore, since the significance of the portions of the signs are often distinct and names for the distinct parts are in common use even for those combined signs that have multiple significance, it would seem better to encode the portions distinctly, just as is done for conjunct consonants in Devanāgarī script. Unless some technical reason contravenes, I would therefore recommend that the following be encoded instead of 1A43-1A45:

1A43 combining virāma (can come after candrabindu and after combining candrabindu)

1A44 combining candrabindu

1A45 combining digit 2 below

xxxx combining digit 3 below

xxxx combining avagraha

It is highly likely that we will find a candrabindu with digit 1 below and candrabindu with digit 4 below given that we have candrabindu above a vowel followed by the digit 1 or the digit 4.

VI. Virāma

If we make candrabindu one character and the sign beneath it (2, 3, avagraha, virāma, second candrabindu) another character, some logic will have to be changed to put virāma beneath it since now *virāma* is not only a character but is also the combination symbol, e.g. t + halant + k yields conjunct consonant. Must have a separate combining virāma character that itself appears rather than signifies that a following character combines.

VII.   Accented visarga

In Table XX Vedic Extensions, we presently slate characters 1A4B, 4C, and 4D Svarita, anudātta, and udātta visarga for encoding. This will not allow separate coloring of the tone sign from the visarga sign. Manuscripts that show tone in red, typically show the boomerangs around the upper and lower dots of the visarga and the horizontal stroke between the dots in red while the two dots themselves are in black. Therefore, the upper boomerang representing the udātta, the lower boomerang representing the anudātta, and the horizontal stroke through the middle representing the svarita must be separate characters from the visarga character. (This last could be handled like the Maitrāyaṇī dependent svarita with the existing Unicode character Stroke Overlay Combining Short (non-connecting) [0335].) Ligate without creating a precomposed glyph so that the accent mark alone can be put in red.

A   White   Yajurvedic   text   called   *Rudrajāpya*   (Poleman   #93, Houghton2007Jan21/MSIndic371/TIFConverts/3v4r.tif) demonstrates this. Text is in black, but the accents are marked in red.

In folio 3 verso line 5, we find a red left-handed boomerang around the lower dot in यादतः ॥ गावो = *yádataḥ . gā́vo*
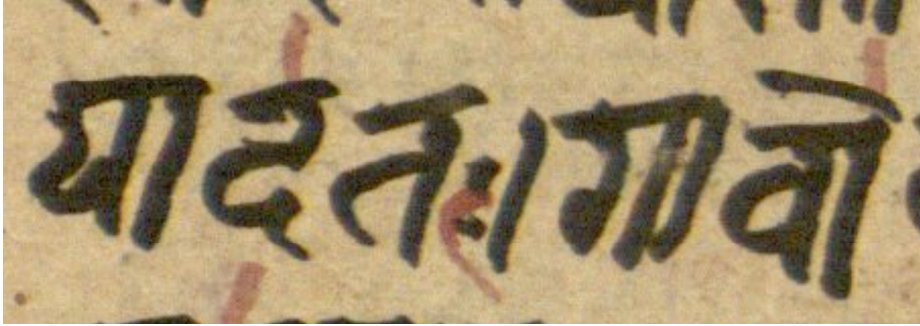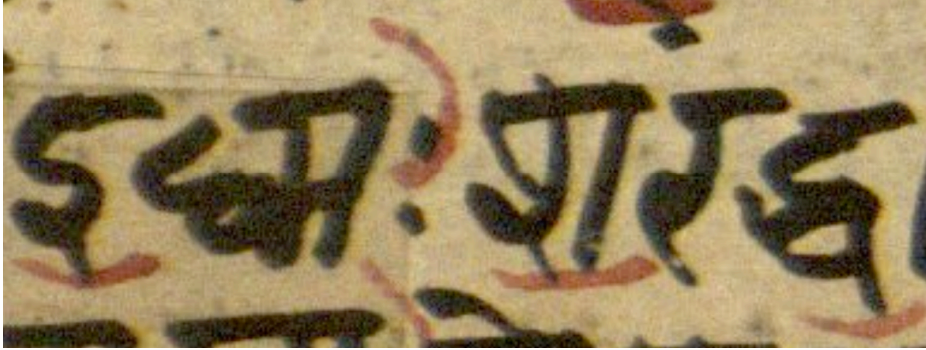

In folio 4 recto line 1, we find a red horizontal stroke between the dots in यद्वैश्यः पद्भ्यां = *yádvaíśyaḥ padbhyā́ṁ*
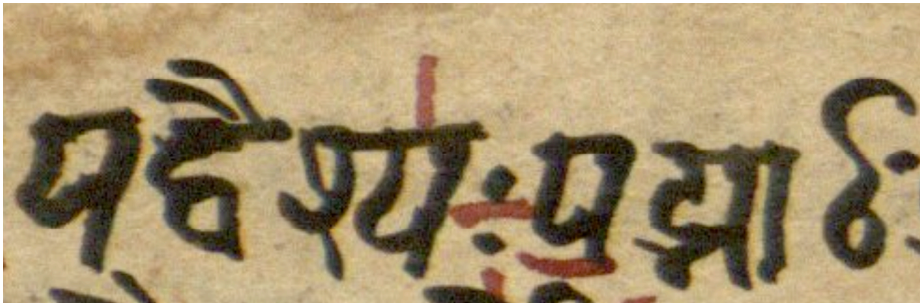
In folio 4 recto line 6, we find a red right-handed boomerang around the upper dot in इध्मः

शरद् = *idhmáḥ śarád*

VIII.   Scans

Many of the scans linked in the documents referenced herein provide better evidence (clearer images with definite textual references) than the ones downloaded from the web that you included on 17 January. Please consider substituting these in the proposal. At present many of the headings are unrelated to the images or other content there.