Comments on L2/08-196 regarding the encoding of Sanskrit and Vedic.

Peter M. Scharf
Thursday, 8 May 2008

The following comments address specific points in the Document L2/08-196, "Proposal for Encoding of Vaidika Sanskrit Characters & Symbols in the BMP of UCS" submitted by the Ministry of Communications and Information Technology Government of India.

1.　Concerning 1CF1 VEDIC SIGN ARDHAVISARGA
=*vaidika jihvaamuliiya upadhamaaniiya*

Document L2/08-196 states, "Experts informed that the usage of Ardhavisarga is before the consonant to which it is applied so the glyph should be modified to ...." precede the consonant to which it is applied. The jihvamuliya and upadhmaniya are allophones of the visarga. Hence it would make most sense to have their behavior align with that of the visarga. All three characters are releases that follow a vowel. While it is true that jihvamuliya precedes only k, kh, and upadhmaniya precedes only p, ph in phonetic order, jihvamuliya and upadhmaniya are not modifications of k, kh, p, ph; rather they are separately recognized sounds that precede these consonants. In certain contexts such as grammatical texts, it would be desirable to refer to the ardhavisarga character separately from the following k, kh, p, ph; hence ardhavisarga should not be made dependent upon these.

2.　Concerning Marks of nasalization 1CE9, 1CEA, 1CEB, 1CEC

The statement in L2/08-196 concludes agreement of the GOI and Scharf/Everson approaches with the tentatively adopted encoding of these four characters at the UTC February meeting. Given this agreement, these four characters should now be adopted into Unicode.

One of the prior objections raised by Professor Joshi to encoding the gomukhas without bindu or candrabindu was that gomukhas without these diacritics were not significant. One idea that now seems feasible on the basis of the current agreement is to encode four gomukha characters with bindu. These would be significant, thereby overcoming Professor Joshi's prior objection. The glyphs with candra bindu over gomukha could be achieved as glyph variants in fonts. There is no difference in the significance between a gomukha with bindu and a gomukha with candrabindu. On the other hand, it is only for Devanagari vowels that the principle has been adopted to encode only significant signs rather than graphic elements. (Compare Eric Muller on prishthamatra in point 4 below.) There is no reason not to analyze the gomukha+bindu into a sequence of two characters.

3.　Concerning

1CEE　VEDIC SIGN HEXIFORM LONG ANUSVARA
1CEF　VEDIC SIGN LONG ANUSVARA
1CF0　VEDIC SIGN RTHANG LONG ANUSVARA

Document 08-196 states, "Code point 1CEF and 1CF0 are representing the same "Gumkar" sign and have been found to be glyph variation in different manuscripts. Therefore, 1CF0 is recommended for encoding and 1CEF will be covered in the write-up as allographic variation and need not be encoded."

Scharf/Everson 07-343 proposed 1CE8 = just the character now shown as 1CEF. As discussed with Professor Joshi in Versailles last October, Scharf/Everson accepted the desire of the GOI to encode these three as separate characters even though 1CEE and 1CF0 are derivates from 1CEF. However, both 1CEE and 1CF0 share the same status as derivative characters, so either both should be withdrawn or both kept on the same principle. They are both typographic imitations of 1CEF that now have obtained distinct recognition and naming conventions. For discussion, see

http://www.unicode.org/~emuller/southasia/vedic/#nasals

For evidence, see Section C1 in the document

http://www.language.brown.edu/Sanskrit/VedicUnicode/passages/VedicMarks2007Mar10V.pdf
(or navigate to VedicMarks2007Mar10V.pdf from
http://sanskritlibrary.org/VedicUnicode/)

and

http://sanskritlibrary.org/VedicUnicode/passages/LongAnusvaraSynonymy.pdf

4.   Pristhamatra

It is correct that in manuscripts the prishthamatra character is generally connected to the FOLLOWING character by a short head stroke unlike the dependent vowel long A which is connected to the PRECEDING character by a short headstroke. However, modern typography obliterates this distinction by making a uniform flat headstroke connecting most characters uninterrupted by space. These characters are not properly Vedic characters since they are common in historical Sanskrit documents regardless of whether they are Vedic or not. They are an early form of Devanagari.

The approach adopted by 08-196 is the approach favored by Eric Muller in his summary of the Vedic discussion on the following grounds, "The model that has been followed so far in Devanagari is to encode each vowel sign separately, even if they can be analyzed graphically as composites."

http://www.unicode.org/~emuller/southasia/vedic/#prishthamatra
Any approach that allows the use of these historically important characters is agreeable to me.

5.   Pushpika

0973 DEVANAGARI SIGN PUSHPIKA
=*vaidika pushpika*
Concerning the statement in 08-196, "The PUSHPIKA is used in Vaidika text only so it should be shifted to the Vaidika extensions block from the proposed Devanagari block. It is used to represent end of text in Vaidika Literature," the sign is found in Devanagari

manuscripts regardless of whether they are of Vedic texts or not so it is not proper to call it Vedic.

6. Caret

0974 DEVANAGARI CARET
=vaidika trutikaa

It does not seem advisable to withdraw this common editorial character on the grounds that there may be other editorial characters. Yet if a proposal of of the other characters can be assembled soon before the end of the comment period, there will be no objection to moving it to join them in a separate block.

7. Heavy ya

097A DEVANAGARI LETTER HEAVY YA

The phonetics of the character are the same as those of the Bengali *09AF*; it is not suggested by the comment, "***used for an affricated glide – 09AF Bengali letter ya***," that the Devanagari character 097A is used in Bengali. Is it used to represent Bengali in Devanagari script? In any case, it is a character proper to Devanagari script. It would not be used to represent Vedic in Telugu, for instance; hence it seems best to call it Devanagari something or other. If it is desired to remove it from the main Devanagari block, let it be put in the Devanagari extended block.

8. Avestan

One book is enough to establish the need for a means to transcribe Avestan into Devanagari. The Avesta is a very significant text and the publication of it in Devanagari transcription demonstrates the need for a digital means of representing the characters used for such a transcription. If in addition a Gujarati character is required, that may be added to the Gujarati block; it does not replace the need for the character in the Devanagari block.

9. Additional Characters

1CF7 Vaidika Anusvaara Ubhayatomukha

The character should be encoded; it appears as

A8F8 DEVANAGARI SIGN ANUSVARA UBHAYATO MUKHA

in N3383R = L2/08-050R. I agree that there is no reason that this couldn't be named and placed in Vedic rather than the Devanagari extension block with the gomukhas or at 1CF1 moving the current 1CF1 to 1CF2.

1CF0 Urasi Rekhaa

The comment, "This may be represented using combination of two code points. Hence there is no need to encode it," is not comprehensible to me. The Mid-char svarita (= Vaidika Svarita Urasi Rekhaa) has been rightly proposed, should be encoded, can only be encoded as a single character, and may be suitably placed at 1CD4. See L2/08-214.

1CF4: Swastika

I concur with the GOI position, even as a Jew.