

Misrepresentation in Unicode of characters related to the Sanskrit sounds Jihvamuliya and Upadhmaniya

Shriramana Sharma – jamadagni-at-gmail-dot-com

2009-Oct-09

OCF1 and OCF2 Kannada Signs Jihvamuliya and Upadhmaniya

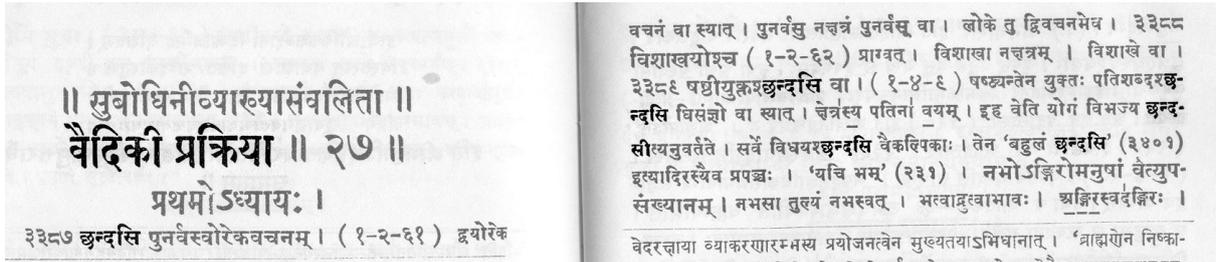
Inappropriateness of labeling as Vedic characters

The Kannada block code chart at the very end has the following text regarding OCF1 KANNADA SIGN JIHVAMULIYA and OCF2 KANNADA SIGN UPADHMANIYA:

VEDIC SIGNS:

The Vedic Signs for jihvamuliya and upadhmaniya were encoded in the Kannada block, but are intended for general Vedic use with all scripts.

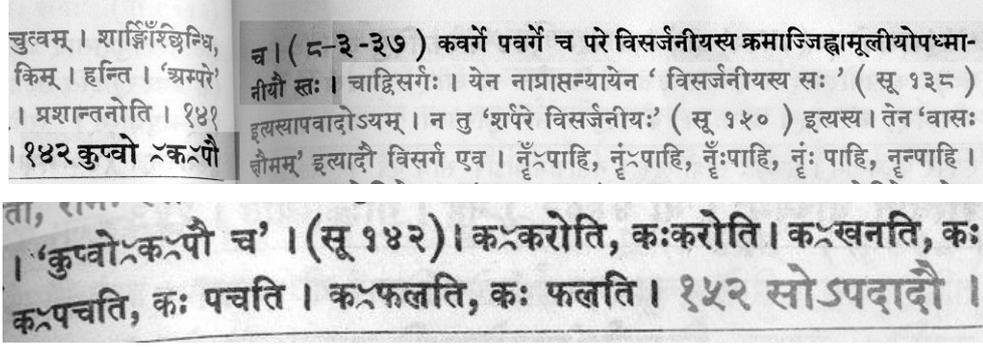
This is incorrect on two counts. First, the sounds jihvāmūliya and upadhmanīya are not specifically “Vedic”. Pāṇini, in his famous grammar work, the Aṣṭādhyāyī, clearly marks Vedic usages by the term “chandasi” (meaning “in the Veda”). This is seen in the following sample from pp 380 and 381 of Siddhānta Kaumudī (a commentary on Pāṇini’s Aṣṭādhyāyī) part 4, reprinted 1997 by Motilal Banarsidass Publishers, Delhi:



Pāṇini has given the rules for the formation of the jihvāmūliya and upadhmanīya as:

kupvo ḥkaḥpau ca (8-3-37)

(As per ISO 15919 ḥ = visarga, ḥ = jihvāmūliya, ḥ = upadhmanīya.) This aphorism is seen in the samples from pp 142, 143 and 147 of part 1 of the same previously-mentioned Siddhānta Kaumudī publication shown at the top of the next page. Translated, this aphorism means “the visarga may be replaced by the jihvāmūliya and upadhmanīya when it is followed by velar and bilabial (voiceless) consonants”.



As it is evident from the above sample, Pāṇini has not mentioned the word “chandasi” in this aphorism, nor has the commentator mentioned it as to be inferred or implied. Thus Pāṇini has not prescribed these solely for Vedic usage. This may be confirmed from any other good commentary or translation of the Aṣṭādhyāyī as well.

Furthermore, scholars of ordinary Sanskrit (and not just specialized Vedic scholars) do pronounce the jihvāmūliya and upadhmanīya in the prescribed contexts as part of normal Sanskrit. This is a ground reality that I as a native speaker of Sanskrit can only state as a fact. I am not aware of any publication that mentions this explicitly which I can provide here as proof of my statement. (Note that by ‘scholars’ I here refer to traditional paṇḍita-s of India who can speak sufficiently fluently in Sanskrit and with whom I regularly interact with as one of the community, and not those who may be quite knowledgeable about the language but cannot speak fluently in the language.)

Therefore it is entirely incorrect to label any characters in any script representing these sounds of Sanskrit as “Vedic” characters which would wrongly imply that they are not used in normal (non-Vedic) Sanskrit.

Inappropriateness of generalizing for pan-Indic use

Further, it is improper to say that these two particular characters encoded in the Kannada block should be used with all scripts. These symbols were encoded in the Kannada block with their particular representative glyphs because Kannada dictionaries show these glyphs being used for these particular characters, as shown in the samples taken from N3366 and shown as figure 1 on the top of the next page.

However, the use of these particular glyphs for the equivalents of these characters in other scripts has not been attested. In fact, Dr Anthony Stone’s well-known pages on ISO 15919 show these glyphs only for Kannada, and quite distinct glyphs for Devanagari as shown in figure 2 taken from <http://homepage.ntlworld.com/stone-catend/trimain1.htm>. Dr Stone has also collected variegated forms of jihvāmūliya and upadhmanīya (see

Presumably the current remarks in the Kannada code chart were placed there because of this request of N3366. However, N3366 has provided no proof that the native users of other Indic scripts recognize these particular glyphs as representative of the jihvāmūliya and upadhmāniya, neither is there at all any proof to provide because this simply is not true.

This is quite unlike the very widespread usage of 0951 and 0952 to denote the svarita and anudātta across Indic scripts which is attested by the very many Vedic books printed in those scripts. Therefore, while it was appropriate of N3366 to request that 0951 and 0952 have their script property changed to common, it was quite inappropriate to request the same of 0CF1 and 0CF2. (This is obviously a late objection, but I cannot help it as I was not aware of N3366 at the time of its submission or approval process.)

Other written forms are not merely glyphic variants

Further, it is not appropriate to say that other forms of jihvāmūliya and upadhmāniya (as shown by Dr Stone in his PDF document mentioned above) may be considered mere glyphic variants. I believe that only those glyphs which are alternatively used for the same character *in the same script context* may be considered variants representing the same character. For example, the two different styles of writing Devanagari A AA etc (as shown at <http://sanskritweb.net/cakram/index.html>, retrieved 2009-Oct-09) may be considered glyphic variants because they are both used in the context of Devanagari only. As a counter-example, the various glyphs across the Indic scripts for the character avagraha may *not* be considered mere glyphic variants of a single avagraha character because they are each limited to their own script context.

Similarly, when the glyphs shown in the Kannada block for jihvāmūliya and upadhmāniya are limited to the Kannada script only, it is *not* appropriate to say that these other forms shown by Dr Stone (chiefly in the context of Devanagari, I may safely add) are mere glyphic variants of the same characters.

Different forms may need to be encoded as different characters

Further, if any requests are made to separately encode distinct forms of the jihvāmūliya and upadhmāniya with sufficient proof for the use of those distinct forms in a particular script (obviously other than Kannada and for example, Devanagari), such requests must not be refused with a recommendation to use the Kannada-specific characters instead.

The problem is here not merely the word ‘Kannada’ in the character names of 0CF1 and 0CF2, but that users of non-Kannada scripts do not recognize these characters as

representative of jihvāmūliya and upadhmāniya and if they have a tradition of representing these two sounds with distinct written forms, they will obviously consider it unacceptable to be forced to use the Kannada-specific forms in plaintext.

Just as avagraha-s are separately encoded for each script because each script has its own consistent use of the avagraha, it must be permitted to encode distinct characters for jihvāmūliya and upadhmāniya if there is sufficient proof of the usage of distinct glyphs for those sounds in other scripts. There is in any case no proof that these Kannada symbols for jihvāmūliya and upadhmāniya are used in other scripts also.

Furthermore, 0CF1 and 0CF2 have GC=So, which quite surprises me. In my understanding, a ‘letter’ is a character that directly denotes a particular sound or sequence of sounds that forms words in a language. There is no reason to not consider these characters (even in Kannada) ‘letters’ by this definition, since they should also be considered parts of words in text operations such as “Select word at cursor” etc. GC=So is appropriate for characters that do not relate to language. In fact, as far as I can see in the entire Unicode 5.2 UnicodeData.txt, there is no character that is used to writing sounds or words of languages that takes GC=So except these two! (I am not counting squared, circled etc versions of Latin or other letters here, since they are not used for writing languages.)

While stability considerations may not permit changing the GC of 0CF1 and 0CF2 despite this being highly desirable to permit text operations to identify them as part of words, at least other characters for jihvāmūliya and upadhmāniya must be able to get GC=Lo. This would be an additional reason to encode separate characters for other scripts.

Request regarding the description in the Kannada code chart

Thus the section heading “Vedic Signs” in the Kannada block code chart must be replaced by “Other signs” or “Additional Signs for Sanskrit” and the following text “The Vedic signs ... all scripts” must be removed. The script property of 0CF1 and 0CF2 must be changed back to ‘Kannada’. Even if this latter request is not possible due to any stability considerations, at least the misleading heading and text in the Kannada code chart must be removed.

1CF2 Vedic Sign Ardhavisarga

1CF2 VEDIC SIGN ARDHAVISARGA was proposed by N3366 as part of a large set of Vedic characters. It has been accepted for Unicode 5.2.0 and has the informative alias “vaidika jihvaamuuliiya upadhmaaniiya”. However, we shall show that this alias should be removed and an annotation should be added to clarify that this character is not limited to Vedic use.

N3366 reads:

VEDIC TONE [sic] ARDHAVISARGA is used to mark either jihvāmūliya (which is a velar fricative [x] occurring only before unvoiced velar stops KA or KHA) or upadhmāniya (which is a bilabial fricative [ɸ] occurring only before unvoiced labial stops PA or PHA).

The samples I have shown in the beginning of this document also clearly show this of use the ardhavisarga to denote the jihvāmūliya and upadhmāniya. As I have already shown that in the first place jihvāmūliya and upadhmāniya cannot be considered as sounds specific to Vedic Sanskrit, it is definitely not appropriate to label any character used for representing these as ‘Vedic’. Thus this character should have been labeled merely “Ardhavisarga” and not “Vedic Sign Ardhavisarga”.

Though it is not now possible to modify the character name, it is possible to remove the misleading informative alias “vaidika jihvaamuuliya upadhmaniya” (“vaidika” is Sanskrit for “Vedic”) and add an annotation clarifying the use of this character.

Request regarding the description in the Vedic Extensions code chart

I request that the informative alias for this character be removed and the following annotations be added instead:

1CF2 VEDIC SIGN ARDHAVISARGA

- * Denotes the sounds jihvamuliya and upadhmaniya (velar and bilabial voiceless fricatives) in Sanskrit.
- * Despite its name and inclusion in this block, this character is not limited to Vedic.

If these annotations are considered too long, they may be truncated to “Denotes the sounds jihvamuliya and upadhmaniya” and “This character is not limited to Vedic” if these would be sufficient.

-o-o-o-