

Request for encoding 1CF3 VEDIC SIGN ROTATED ARDHAVISARGA

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

2010-Jul-09

Attention: This document supersedes my previous submission

L2/09-343 Request for encoding 1CF3 ROTATED ARDHAVISARGA dated 2009-Oct-09.

This is a request for encoding a character in the Vedic Extensions block. This character resembles the 1CF2 VEDIC SIGN ARDHAVISARGA, however it is rotated through 90° so that it more resembles a vertical biconcave lens:



1CF2 VEDIC SIGN ARDHAVISARGA



*1CF3 VEDIC SIGN ROTATED ARDHAVISARGA

This character has been used in contrast with 1CF2 VEDIC SIGN ARDHAVISARGA in some contexts, as in the ALA-LC romanization table for Sanskrit and Prakrit (ref 1):

Anunāsika

ँ ढ

Visarga

: ः

Jihvāmūlīya

) (ॡ

Upadhmāñīya

⌘ ॢ

Avagraha

(see Note 4)

₡ , (apostrophe)

and in the collection made by Dr Anthony Stone of different glyphs he has noticed attested for the distinct representation of the jihvamuliya and upadhmaniya (ref 2):

<i>Journal of the Bihar and Orissa Research Society,</i> transliteration table	✕) (
Anshuman Pandey, devchars.pdf	⌘) (
<i>ALA-LC Romanization Tables, 1997, 'Sanskrit and Prakrit'</i>) (⌘

Since this character is used contrastively with 1CF2 VEDIC SIGN ARDHAVISARGA in these contexts, separately encoding this character is justifiable.

Those who would like to follow the ALA-LC–attested distinction between these two characters in plaintext would be benefited to have a separate encoding for the rotated form of the ardhavisarga compared to the regular ardhavisarga at 1CF2.

Justification for encoding such a rare-use character

I accept that there is no widespread usage of this character, but its occasional usage in contrast with the currently encoded ardhavisarga is nevertheless attested by respectable researchers such as those of the ALA-LC and Dr Anthony Stone. I believe that students of Sanskrit phonetics – at least those in the traditional scholarly community of India which still thrives in places, and who often have insufficient knowledge of or exposure to English to find the IPA comfortable for their needs – will be benefited from being able to use this character in the same plaintext as and distinct from the currently encoded ardhavisarga.

Therefore I believe this character is a valid candidate for encoding and hence request the UTC to encode it at 1CF3 next to the regular ardhavisarga.

Whether to encode as combining mark or not

I have previously (via postings on the unicode@unicode.org mailing list in 2009-Aug-21 and following) brought to the attention of the Unicode community some annoying problems encountered while composing Sanskrit texts caused by the specification of Mc as the GC of the visarga, anusvara and related characters in Indic scripts (which includes ardhavisarga).

I also point out that Dr Peter Scharf whom the UTC has contacted w.r.t my submission also said in his mail to Deborah Anderson and me dated 2009-Nov-01 00:48 (UTC+0530):

I fully agree with Sharma that none of these characters (jihvāmūliya, upadhmānīya, ardhavisarga, or rotated ardhavisarga) should be combining characters or modifying characters; they are independent signs for consonantal sounds. Only their line-breaking properties need be restricted.

However, the UTC does not wish to make such a broad-spread change without considering the potential implications on the current data containing these characters. Until such a document may be submitted as discusses those implications, it would be appropriate to maintain status quo and encode these characters with GC=Mc (or GC=Mn).

Choice of character name and block

In my previously-mentioned document L2/09-342 regarding the jihvamuliya and upadhmaniya, I have requested an annotation to be added to 1CF2 to clarify that it is not limited to Vedic use despite its name of “Vedic Sign Ardhavisarga”. I have provided proof for this in that document. The same proof applies for the character proposed by the present document also. Since it is considered better to label this newly proposed character with the adjective VEDIC SIGN in its character name to maintain uniformity with 1CF2 VEDIC SIGN ARDHAIVISARGA, this newly proposed character should also carry such an annotation. Instead of duplicating the annotation for each character, the editors may consider putting these characters under a sub-heading in the codechart and adding these remarks under the heading just as has been done for the generic Indic danda-s in the Devanagari block.

As regards the name proper, I propose “Rotated Ardhavisarga”. The adjective describes the orientation of this character compared to the normal ardhavisarga. The noun is also directly appropriate for this character since from the attestation provided by Dr Stone (ref 2) it is evident that this character has been used for both the jihvamuliya and the upadhmaniya, just like the regular ardhavisarga.

Despite this character not being limited to Vedic use I ask for it to be placed in the Vedic Extensions block only to ensure visibility by being near 1CF2.

Technical Aspects

The Unicode character properties for this character would be:

1CF3;VEDIC SIGN ROTATED ARDHAIVISARGA;Mc;0;L;;;;N;;;;;

As for collation, the jihvamuliya is normally ordered in Sanskrit before the upadhmaniya, and hence wherever 1CF2 VEDIC SIGN ARDHAIVISARGA or 1CF3 VEDIC SIGN ROTATED ARDHAIVISARGA are used, some minor contextual analysis may need to be done to determine which sound is represented. Obviously, when a velar consonant follows, it is the jihvamuliya and when a bilabial consonant follows, it is the upadhmaniya. Ordering must be done accordingly.

As for linebreaking, 1CF2 definitely will not take a linebreak before, since it is a combining mark. 1CF3 should not either. Both should also not take a linebreak after, since their semantic content sometimes depends on the nature of the following consonant and hence native users do not split lines between an ardhavisarga and the following consonant.

In the code chart, annotations should be added to this character as follows:

1CF3 VEDIC SIGN ROTATED ARDHAIVISARGA

- * Denotes the sounds jihvamuliya and upadhmaniya (velar and bilabial voiceless fricatives) in Sanskrit
- * Sometimes used in contrast with 1CF2 VEDIC SIGN ARDHAIVISARGA with each representing one of the two possible sounds.
- * Despite being encoded 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”, “Sometimes used in contrast with 1CF2 VEDIC SIGN ARDHAIVISARGA” and “This character is not limited to Vedic” if these would be sufficient.

References

1. ALA-LC Romanization for Sanskrit and Prakrit, Library of Congress, 1997, retrieved from <http://www.loc.gov/catdir/cpsa/romanization/sanskrit.pdf> on 2009-Oct-09. ALA-LC Romanization Tables page: <http://www.loc.gov/catdir/cpsa/roman.html>
2. Some usages of written forms of the Jihvamuliya and Upadhmaniya, Dr Anthony P Stone, retrieved from <http://homepage.ntlworld.com/stone-catend/jihvupadh.pdf> on 2009-Oct-09.

-0-

Official Proposal Summary Form

A. Administrative

1. Title

Proposal for encoding 1CF3 VEDIC SIGN ROTATED ARDHAVISARGA

2. Requester's name

Shriramana Sharma

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution

4. Submission date

2009-Oct-13 (original), 2010-Jul-09 (revised, now)

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes.

1d. Name of the existing block

Vedic Extensions

2. Number of characters in proposal

1 (one)

3. Proposed category (A-Contemporary)

Category A.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Shriramana Sharma (for this character alone).

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Shriramana Sharma, FontForge. TTF file submitted alongwith proposal.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.

See detailed proposal.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Dr Mani Dravid, Madras Sanskrit College, Chennai

2c. If YES, available relevant documents

None specifically. Mode of contact was personal conversation.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Vedic and Sanskrit scholars.

4a. The context of use for the proposed characters (type of use; common or rare)

Somewhat rare in Vedic and Sanskrit scholarly contexts.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes, occasionally.

5b. If YES, where?

In Sanskrit scholarly communities in India and elsewhere.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

The rationale is that related characters are encoded in the BMP and there is space in the block.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Only one character is proposed. It should be placed immediately after its relative, 1CF2.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

It resembles the character 1CF2 VEDIC SIGN ARDHAVISARGA rotated through 90°.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

The rationale is that it has been used and needs to be used in plaintext in contrast with 1CF2 VEDIC SIGN ARDHAVISARGA. This is illustrated in the detailed proposal.

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

-O-O-O-