

Title: Proposal to Encode the SANDHI MARK for Sharada
Source: Script Encoding Initiative (SEI)
Author: Anshuman Pandey (pandey@umich.edu)
Status: Liaison Contribution
Action: For consideration by UTC and WG2
Date: 2012-09-27

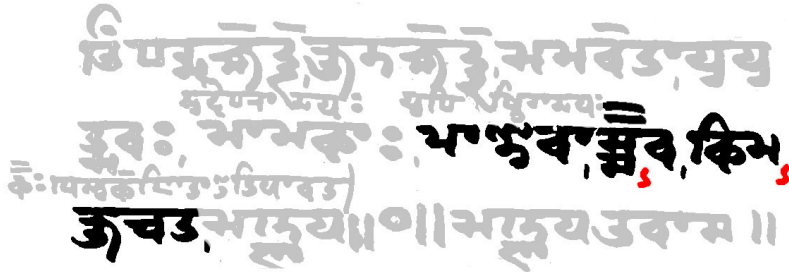
1 Introduction

This is a preliminary proposal to encode a new character in the Sharada block of the Universal Character Set (ISO/IEC 10646). Properties of the proposed character are:

| GLYPH | CODE | CHARACTER NAME | GC | CCC | BIDI | MIRRORED | LINEBREAK |
|-------|-------|---------------------|----|-----|------|----------|-----------|
| ₵ | 111C9 | SHARADA SANDHI MARK | Po | 0 | L | N | AL |

2 Description

The ₵ SHARADA SANDHI MARK is used in some Sharada manuscripts for indicating external sandhi in Sanskrit documents. It is a below-base mark that is written after the syllable where sandhi occurs. The excerpt below, from *Bhagavad Gītā* 1.1, illustrates usage of the mark (adapted from Lokesh Chandra 1982):



Here the SANDHI MARK occurs in the phrases *pāṇḍavāścaiva* and *kimaḥkurvata*. In the first phrase, it indicates vowel sandhi between *ca* and *eva* ($^{\circ}a e^{\circ} \rightarrow ^{\circ}ai^{\circ}$); in the second it indicates fusion of a bare consonant and a vowel between *kim* and *akurvata* ($^{\circ}m a^{\circ} \rightarrow ^{\circ}ma^{\circ}$).

In this manuscript of the *BG*, the SANDHI MARK is written twice in order to indicate sandhi between two long vowels, as in the following excerpt from *BG* 2.1:



Here the doubled mark occurs in the phrase *kṛpayāviṣṭam*, where it indicates sandhi of identical long vowels between *kṛpayā* and *āviṣṭam* ($^{\circ}\bar{a} \bar{a}^{\circ} \rightarrow ^{\circ}\bar{a}^{\circ}$). However, the double SANDHI MARK is not consistently throughout the manuscript for marking long-vowel sandhi.

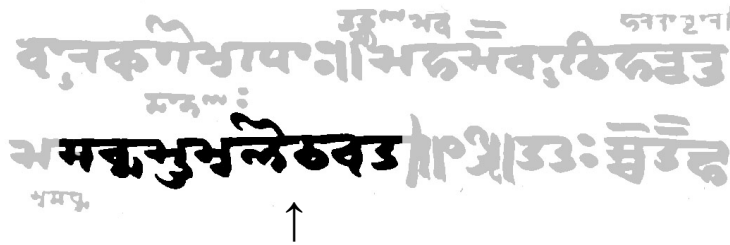
3 Relationship to Avagraha

The graphical shape of , SHARADA SANDHI MARK is similar to that of ; U+111C1 SHARADA SIGN AVAGRAHA. To be sure, the AVAGRAHA is also a type of sandhi mark, but it is used specifically in Sharada and other Indic scripts to indicate the elision of word-initial *a* that results from sandhi initiated by the final syllable of the preceding word. In the Sharada manuscript of the *Bhagavad Gītā* that is analyzed here, the SANDHI MARK and AVAGRAHA are distinctive, as is illustrated in the following example:

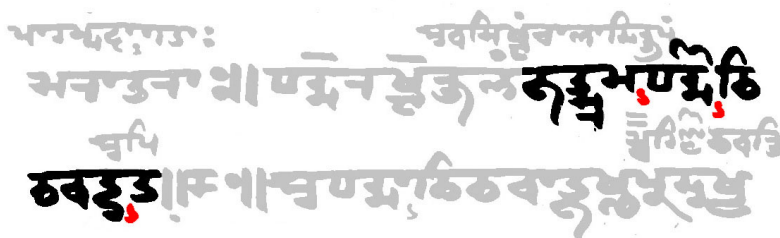


Here, the AVAGRAHA (transliterated as apostrophe) occurs in BG 2.22 in the phrase *naro'parāṇi* to indicate *a*-elision in *aparāṇi* on account of the *visarga* in *naro* (*naraḥ*), following the rule ($^{\circ}aḥ a^{\circ} \rightarrow ^{\circ}o' ^{\circ}$).

However, the usage of AVAGRAHA is just as inconsistent as usage of SANDHI MARK in this manuscript. In the below excerpt from verse 1.13, an AVAGRAHA would be expected at the position marked by the arrow, between *śabdastumulo* and *abhavat*:



In other parts of the text, the SANDHI MARK acts both as an AVAGRAHA and as a general indicator of external sandhi. The following excerpt from BG 1.40 shows three occurrences of the SANDHI MARK:



In the above phrase *kṛtsnamadharmobhivhavatyuta*, the first and third SANDHI MARK indicate external sandhi between *kṛtsnam* and *adharmo* ($^{\circ}m a^{\circ} \rightarrow ^{\circ}ma^{\circ}$) and between *abhivhavati* and *uta* ($^{\circ}i u^{\circ} \rightarrow ^{\circ}yu^{\circ}$), respectively. The second mark plays the role of AVAGRAHA and represents elision of the initial *a* in *abhivhavati* on account of the *visarga* in *adharmo* (*adharmah*).

Given the function of AVAGRAHA, it is highly likely that usage of the character was expanded by Sharada scribes for indicating other types of sandhi, resulting in the subscript form. While the subscript *avagraha* may be used by some scribes to represent both external sandhi and *a*-elision, the normative AVAGRAHA cannot be used generally for the former. For this reason, it is necessary to encode the SANDHI MARK as an independent character.

4 Encoded Representation

The SANDHI MARK is written after an orthographic syllable in order to indicate that the syllable represents a word boundary that is affected by sandhi. The mark is not intended for specifying linguistic syllables and, therefore, it does not delimit correct morphological boundaries. This is explained below. In the following excerpt from BG 1.21, the text *vākyamīdamāha* is highlighted:

The mark occurs after the syllables *mi* and *mā*; however, the true word boundaries of the phrase are *vākyam idam āha*. In the below example from BG 1.40, the text *bhavatyuta* is highlighted:

The mark occurs after the syllable *tyu*; however, the true word boundaries are *bhavaty uta*, or rather *bhavati + uta* ($^{\circ}i u^{\circ} \rightarrow ^{\circ}yu^{\circ}$). In the next example, the text *kṛpayāviṣṭam* is highlighted:

Here, the doubled SANDHI MARK occurs after *yā*, but the true word boundaries are *kṛpayā āviṣṭam*.

The practice of writing the SANDHI MARK after the syllable *tha* embodies the sandhi operation is to be followed in the encoded representation. Following this, the four syllables that occur in the above examples would be represented as:

- मि_₃ (*mi*_₃) = <MA, VOWEL SIGN I, SANDHI MARK>
- भा_₃ (*mā*_₃) = <MA, VOWEL SIGN AA, SANDHI MARK>.
- त्रु_₃ (*tyu*_₃) = <TA, VIRAMA, YA, VOWEL SIGN U, SANDHI MARK>
- घा_{₃₃} (*yā*_{₃₃}) = <YA, VOWEL SIGN AA, SANDHI MARK, SANDHI MARK>.

5 Ordering

The SANDHI MARK may occur together with , U+111C8 SHARADA SEPARATOR and may either precede or follow the separator:



In such cases, the visual order should be preserved in the encoded representation.

6 Glyph Positioning

When the SANDHI MARK is written between a consonant and a left-spacing or non-spacing vowel sign, it should be positioned equidistant between the consonant and the following letter or sign. When it occurs between a consonant and a right-spacing vowel sign, it should be positioned directly beneath the vowel sign. Both of these cases are illustrated below:



When two SANDHI MARK-S occur after VOWEL SIGN A, they are centered beneath the sign.



7 References

Lokesh Chandra [comp.]. 1982. *Sanskrit Texts from Kashmir*, vol. 1. Śata-Piṭaka Series: Indo-Asian Literatures, volume 298. New Delhi: Sharada Rani.

Pandey, Anshuman. 2009. "Proposal to Encode the Sharada Script in ISO/IEC 10646". N3595 L2/09-074R2. August 5, 2009. <http://std.dkuug.dk/JTC1/SC2/WG2/docs/n3595.pdf>

8 Acknowledgments

I would like to thank Shriramana Sharma for his discussions on the proposed character.

This project was made possible in part by a grant from the United States National Endowment for the Humanities, which funded the Universal Scripts Project (part of the Script Encoding Initiative at the University of California, Berkeley). Any views, findings, conclusions or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.

**ISO/IEC JTC 1/SC 2/WG 2
PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646¹**

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://www.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://www.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

| | |
|--|--|
| 1. Title: | Proposal to Encode the SANDHI MARK for Sharada |
| 2. Requester's name: | <i>Script Encoding Initiative (SEI) / Anshuman Pandey (pandey@umich.edu)</i> |
| 3. Requester type (Member body/Liaison/Individual contribution): | <i>Liaison contribution</i> |
| 4. Submission date: | <i>2012-09-27</i> |
| 5. Requester's reference (if applicable): | |
| 6. Choose one of the following: | |
| This is a complete proposal: | <input checked="" type="checkbox"/> Yes |
| (or) More information will be provided later: | <input type="checkbox"/> |

B. Technical – General

| | |
|---|--|
| 1. Choose one of the following: | |
| a. This proposal is for a new script (set of characters): | |
| Proposed name of script: | |
| b. The proposal is for addition of character(s) to an existing block: | <input checked="" type="checkbox"/> Yes |
| Name of the existing block: | <i>Sharada</i> |
| 2. Number of characters in proposal: | <input checked="" type="checkbox"/> 1 |
| 3. Proposed category (select one from below - see section 2.2 of P&P document): | |
| A-Contemporary <input checked="" type="checkbox"/> B.1-Specialized (small collection) <input type="checkbox"/> B.2-Specialized (large collection) <input type="checkbox"/> | |
| C-Major extinct <input type="checkbox"/> D-Attested extinct <input type="checkbox"/> E-Minor extinct <input type="checkbox"/> | |
| F-Archaic Hieroglyphic or Ideographic <input type="checkbox"/> G-Obscure or questionable usage symbols <input type="checkbox"/> | |
| 4. Is a repertoire including character names provided? | <input checked="" type="checkbox"/> Yes |
| a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document? | <input checked="" type="checkbox"/> Yes |
| b. Are the character shapes attached in a legible form suitable for review? | <input checked="" type="checkbox"/> Yes |
| 5. Fonts related: | |
| a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard? | <i>Anshuman Pandey</i> |
| b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.): | <i>Anshuman Pandey (pandey@umich.edu)</i> |
| 6. References: | |
| a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided? | <input checked="" type="checkbox"/> Yes |
| b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached? | <input checked="" type="checkbox"/> Yes |
| 7. Special encoding issues: | |
| 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)? | <input checked="" type="checkbox"/> Yes |

8. Additional Information:

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. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

¹ Form number: N4102-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

C. Technical - Justification

| | |
|--|--------|
| 1. Has this proposal for addition of character(s) been submitted before? If YES explain | No |
| 2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)? If YES, with whom? If YES, available relevant documents: | No |
| 3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: | Yes |
| 4. The context of use for the proposed characters (type of use; common or rare) Reference: | Common |
| 5. Are the proposed characters in current use by the user community? If YES, where? Reference: | Yes |
| 6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP? If YES, is a rationale provided? If YES, reference: | N/A |
| 7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)? | N/A |
| 8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence? If YES, is a rationale for its inclusion provided? If YES, reference: | No |
| 9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters? If YES, is a rationale for its inclusion provided? If YES, reference: | No |
| 10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character? If YES, is a rationale for its inclusion provided? If YES, reference: | No |
| 11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? If YES, reference: Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? If YES, reference: | No |
| 12. Does the proposal contain characters with any special properties such as control function or similar semantics? If YES, describe in detail (include attachment if necessary) | No |
| 13. Does the proposal contain any Ideographic compatibility characters? If YES, are the equivalent corresponding unified ideographic characters identified? If YES, reference: | No |