

Proposal to encode the SANDHI MARK for Bengali

Srinidhi A and Sridatta A
Tumakuru, Karnataka, India
srinidhi.pinkpetals24@gmail.com, sridatta.jamadagni@gmail.com

November 1, 2016

1 Introduction

This is a proposal to encode a new character in the Bengali block. The details of the character are given below.

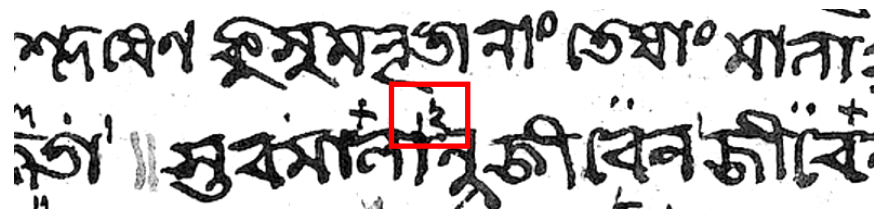
Glyph Code Character name

২

09FE BENGALI SANDHI MARK

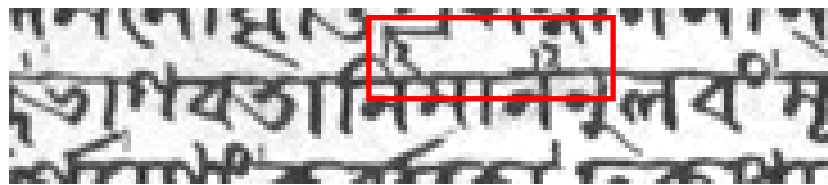
2 Description

The sandhi mark is used in many Sanskrit manuscripts of Bengali script to show external Sandhi. The following section from the manuscript *Stavamāla* demonstrates the use of the sandhi mark.



In this excerpt the sandhi mark occurs in the phrase *Stābamālānujībena*. Here vowel sandhi occurs between the words *Stābamālā* and *anujībena* ($\bar{a}+a \rightarrow \bar{a}$).

Sandhi mark is also sometimes used to indicate the fusion of a bare consonant and a vowel. An example is shown in following manuscript of *Bhāgavatāṣṭakam* between the words *bhāgabatān*, *imān* and *anulabaṃ*.

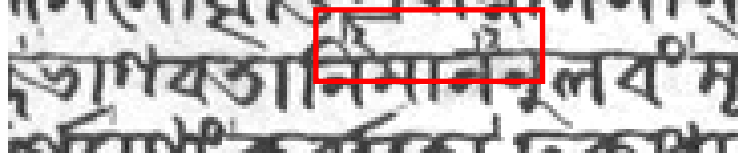


3 Representation

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 sandhi mark is closely related to 111C9 SHARADA SANDHI MARK in its behavior and function. Similar to Sharada, in Bengali too the sandhi mark is not intended for specifying linguistic syllables.

Therefore, it does not delimit correct morphological boundaries. This is elucidated in below manuscript.



Here, sandhi mark is used after ni and na. However the actual linguistic word boundaries of the phrase are bhāgabatān imān anulabaṃ. The sandhi marks should be positioned equidistant between the two surrounding syllables.

4 Character Data

Properties of the proposed character are:

09FE; BENGALI SANDHI MARK;Po;O;L;;;;N;;;;;

Line Break: AL

The character has to be annotated in the Code chart as given below

*i indicates external sandhi in Sanskrit documents

5 Attestations

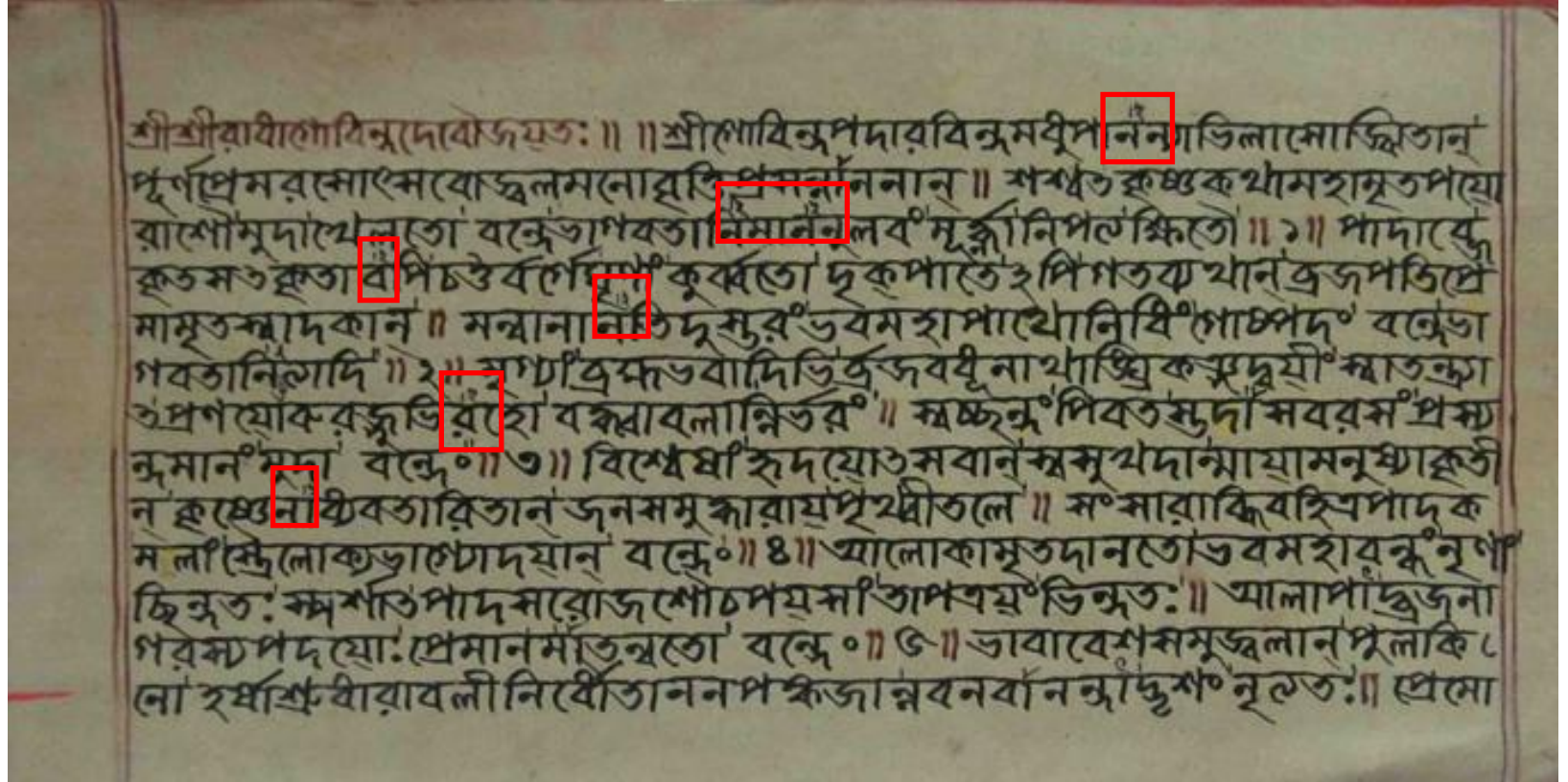


Figure 1. A manuscript from *Bhāgavatāṣṭakam* showing the use of sandhi mark.

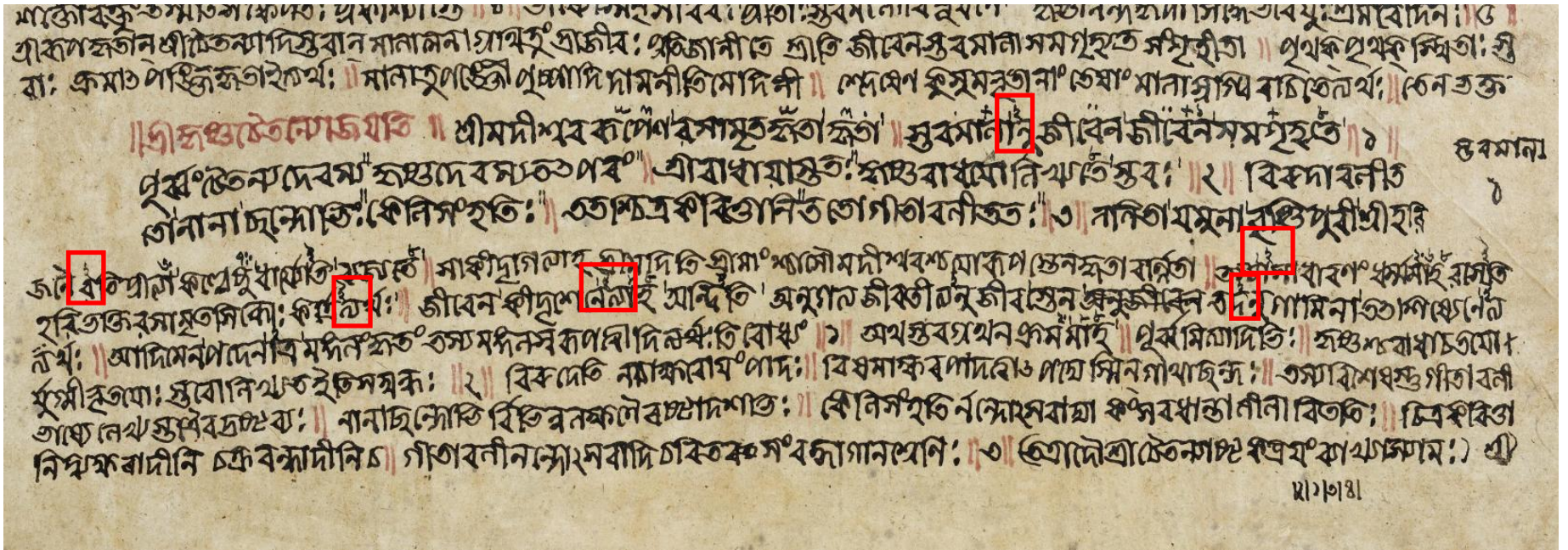


Figure 2. A manuscript from Stābamālā showing the use of sandhi mark.(from ¹)



Figure 3. Use of Sandhi mark in manuscript of Brahmasaṃhitā.

The German text says *avagraha-artiges Element oberhalb des Zeichens: Sandhitrenner*. The text translates Avagraha-like element above the sign used to denote Sandhi (from IndoSript project)

¹ <http://idb.ub.uni-tuebingen.de/diglit/Mal211?sid=24b064cf9bcf0c9fe59a37722e9f75b1>

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS

FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646³.

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

Please read Principles and Procedures Document (P & P) from <http://std.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://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

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

A. Administrative

1. Title:	<i>Proposal to encode the SANDHI MARK for Bengali</i>
2. Requester's name:	<i>Srinidhi A and Sridatta A</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>November 1, 2016</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(or) More information will be provided later:	

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:	<i>Yes</i>	
Name of the existing block:	<i>Bengali</i>	
2. Number of characters in proposal:	<i>1</i>	
3. Proposed category (select one from below - see section 2.2 of P&P document):		
A-Contemporary <input type="checkbox"/>	B.1-Specialized (small collection) <input type="checkbox"/>	B.2-Specialized (large collection) <input checked="" type="checkbox"/>
C-Major extinct <input type="checkbox"/>	D-Attested extinct <input type="checkbox"/>	E-Minor extinct <input type="checkbox"/>

³ Form number: N4502-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)

F-Archaic Hieroglyphic or Ideographic

G-Obscure or questionable usage symbols

4. Is a repertoire including character names provided?

Yes

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

Yes

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

Yes

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

Srinidhi A

b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):

Srinidhi A(Srinidhi.pinkpetals24@gmail.com)

6. References:

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

Yes

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

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)?

Yes

See text of proposal

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.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<i>No</i>
If YES explain	
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.)?	<i>No</i>
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	<i>Yes</i>
Reference:	<i>Used by scholars desiring to document Bengali manuscripts.</i>
4. The context of use for the proposed characters (type of use; common or rare)	<i>common</i>
Reference:	<i>The character is used in some records for indicating sandhi breaks</i>
5. Are the proposed characters in current use by the user community?	<i>Yes</i>
If YES, where? Reference:	<i>Used by specialists studying the Bengali script</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>Yes</i>
If YES, is a rationale provided?	<i>Yes</i>
If YES, reference:	<i>The Bengali block is BMP</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>N/A</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>No</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	<i>No</i>
If YES, is a rationale for its inclusion provided?	
If YES, reference:	

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?

No

If YES, is a rationale for its inclusion provided?

If YES, reference:

11. Does the proposal include use of combining characters and/or use of composite sequences?

No

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:

12. Does the proposal contain characters with any special properties such as

control function or similar semantics?

No

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

13. Does the proposal contain any Ideographic compatibility characters?

No

If YES, are the equivalent corresponding unified ideographic characters identified?

If YES, reference: