

Proposal to encode Bengali Sign Combining Anusvara Above



Jan Kučera, Charles University
Charles Li, University of Hamburg

2024-05-23

This is a proposal to encode 1 character, U+0984 BENGALI SIGN COMBINING ANUSVARA ABOVE.

Like with Newa script's ° U+11444 NEWA SIGN ANUSVARA and ˘ U+11448 NEWA SIGN FINAL ANUSVARA, Bengali and Assamese historical sources also exhibit different kinds of anusvara in a contrastive way.

A **top, non-spacing** form of anusvara can be found in manuscripts sources. In some cases, it is a separate circle above a consonant or vowel, in others it is part of the headline, see the summary in Figure 1 and evidence below.

4. ANUSVĀRA, VISARGA, AVAGRAHA AND OTHER SYMBOLS

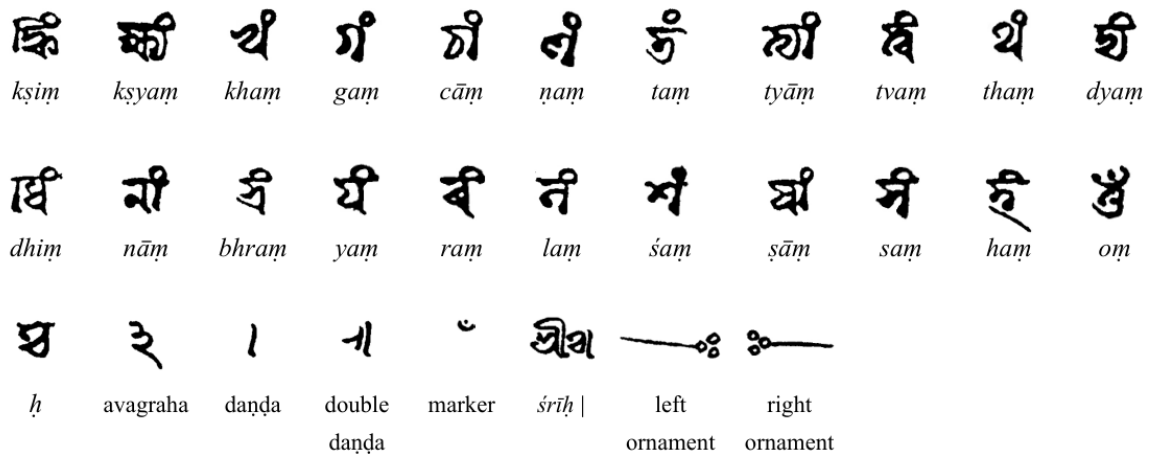


Figure 1. Anusvaras as found in a manuscript of Daṇḍin's Kāvyaḍarśa.

Source: DIMITROV Dragomir. Tables of the Old Bengali Script. In: Dragomir DIMITROV, Ulrike ROESLER, Roland STEINER (ed.). शिखिसमुच्चयः Indian and Tibetan Studies (Collectanea Marpurgensia Indologica et Tibetologica), p. 66. Wien: 2002.

Evidence from manuscripts:



Figure 2. Non-spacing anusvaras on consonants.

Transliteration: Left: *manaso bhīpsitaṃ*... Right: *mūrttidhyāyaṃ*

Source: Garuḍapurāṇa

https://tst-project.github.io/mss/Sanskrit_1419.xml?facs=6



Figure 3. Non-spacing anusvaras on consonants.

Transliteration: Left: *...saṃśrṣṭī*... Right: *svasaṃkrāntadhane*

Source: Saṃkrāntadhanaviṣaya

https://tst-project.github.io/mss/Sanskrit_1442_7a.xml?facs=4

Contrastive evidence with currently encoded anusvara:

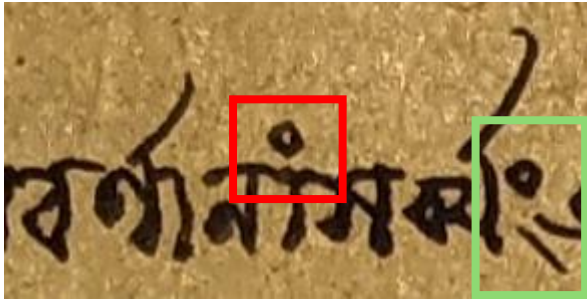


Figure 4. Non-spacing anusvara on vowel sign contrasting with encoded anusvara.

Transliteration: *varṇānāṃ sarvvaṃ*

Source: Mahābhārata — Mokṣadharmaparvan

https://tst-project.github.io/mss/Sanskrit_1442_7b.xml?facs=6

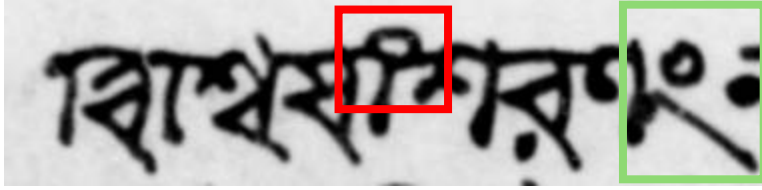


Figure 5. Non-spacing anusvara on vowel sign contrasting with encoded anusvara.

Transliteration: *viśveṣāṃ śaraṃ*

Source: Vivādārṇavasetu

<https://gallica.bnf.fr/ark:/12148/btv1b100915605/f132>

Character properties

UnicodeData.txt

```
0984;BENGALI SIGN COMBINING ANUSVARA ABOVE;Mn;0;NSM;;;;;N;;;;;
```

NamesList.txt

```
0984 BENGALI SIGN COMBINING ANUSVARA ABOVE
```

IndicSyllabicCategory.txt

```
0984; Bindu # Mn BENGALI SIGN COMBINING ANUSVARA ABOVE
```

IndicPositionalCategory.txt

```
0984; Top # Mn BENGALI SIGN COMBINING ANUSVARA ABOVE
```

LineBreaking.txt

```
0984; CM # Mn BENGALI SIGN COMBINING ANUSVARA ABOVE
```

The codepoint and name was chosen to match U+0C04 TELUGU SIGN COMBINING ANUSVARA ABOVE. Alternatively, U+0BFF is available near other signs.

Chart

	098	099	09A	09B	09C	09D	09E	09F
0	৭ 0980	ঐ 0990	ঔ 09A0	ৱ 09B0	ী 09C0		ঋ 09E0	ঌ 09F0
1	ঁ 0981		ড 09A1		় 09C1		ঞ 09E1	ঝ 09F1
2	ং 0982		ঢ 09A2	ণ 09B2	্ 09C2		ৎ 09E2	ঠ 09F2
3	ঃ 0983	ও 0993	এ 09A3		় 09C3		ঔ 09E3	ট 09F3
4	ঁ 0984	ঔ 0994	তা 09A4		় 09C4			ঠ 09F4
5	অ 0985	ক 0995	খ 09A5					ৎ 09F5
6	ব্রা 0986	খ 0996	দ 09A6	শ 09B6			০ 09E6	ঞ 09F6
7	ভ 0987	গ 0997	ঘ 09A7	ষ 09B7	ে 09C7	ী 09D7	১ 09E7	। 09F7
8	ক্ষ 0988	ষ 0998	ন 09A8	স 09B8	ৈ 09C8		২ 09E8	৳ 09F8
9	ডে 0989	ঙ 0999		ত 09B9			৩ 09E9	০ 09F9
A	ঢে 098A	চ 099A	প 09AA				৪ 09EA	৵ 09FA
B	ক্ষ 098B	ছ 099B	ফ 09AB		ৌ 09CB		৫ 09EB	৶ 09FB
C	ঙ 098C	জ 099C	ব 09AC	় 09BC	ৌ 09CC	ড় 09DC	৬ 09EC	৷ 09FC
D		ঝ 099D	ভ 09AD	ৎ 09BD	্ 09CD	ঢ 09DD	৭ 09ED	৸ 09FD
E		ঞ 099E	য 09AE	া 09BE	ে 09CE		৮ 09EE	ঞ 09FE
F	এ 098F	ট 099F	য 09AF	ি 09BF		য় 09DF	৯ 09EF	

The Bengali script is also known as Bangla. In Assam, the preferred name of the script is Asamiya or Assamese. The Assamese language has also been written historically using distinct regional scripts known as Kamrupi.

Various signs

- 0980 ৭ BENGALI ANJI
= siddham, siddhirastu
• used at the beginning of texts as an invocation
- 0981 ঁ BENGALI SIGN CANDRABINDU
- 0982 ং BENGALI SIGN ANUSVARA
- 0983 ঃ BENGALI SIGN VISARGA
- 0984 ঄ BENGALI SIGN COMBINING ANUSVARA ABOVE

Independent vowels

- 0985 অ BENGALI LETTER A
- 0986 আ BENGALI LETTER AA
- 0987 ঐ BENGALI LETTER I
- 0988 ঔ BENGALI LETTER II
- 0989 উ BENGALI LETTER U
- 098A উ BENGALI LETTER UU
- 098B ঋ BENGALI LETTER VOCALIC R
- 098C ঌ BENGALI LETTER VOCALIC L
- 098D <reserved>
- 098E <reserved>
- 098F এ BENGALI LETTER E
- 0990 ঐ BENGALI LETTER AI
- 0991 <reserved>
- 0992 <reserved>
- 0993 ও BENGALI LETTER O
- 0994 ঔ BENGALI LETTER AU

Consonants

- 0995 ক BENGALI LETTER KA
- 0996 খ BENGALI LETTER KHA
- 0997 গ BENGALI LETTER GA
- 0998 ঘ BENGALI LETTER GHA
- 0999 ঙ BENGALI LETTER NGA
- 099A চ BENGALI LETTER CA
- 099B ছ BENGALI LETTER CHA
- 099C জ BENGALI LETTER JA
- 099D ঝ BENGALI LETTER JHA
- 099E ঞ BENGALI LETTER NYA
- 099F ট BENGALI LETTER TTA
- 09A0 ঠ BENGALI LETTER TTHA
- 09A1 ড BENGALI LETTER DDA
- 09A2 ঢ BENGALI LETTER DDHA
- 09A3 ণ BENGALI LETTER NNA
- 09A4 ত BENGALI LETTER TA
- 09A5 থ BENGALI LETTER THA
- 09A6 দ BENGALI LETTER DA
- 09A7 ধ BENGALI LETTER DHA
- 09A8 ন BENGALI LETTER NA
- 09A9 <reserved>
- 09AA প BENGALI LETTER PA
- 09AB ফ BENGALI LETTER PHA
- 09AC ব BENGALI LETTER BA
= Bengali va, wa
- 09AD ভ BENGALI LETTER BHA
- 09AE ম BENGALI LETTER MA
- 09AF য BENGALI LETTER YA
- 09B0 র BENGALI LETTER RA
- 09B1 <reserved>

**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://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>Bengali Sign Combining Anusvara Above</i>		
2. Requester's name:	<i>Jan Kučera, Charles Li</i>		
3. Requester type (Member body/Liaison/Individual contribution):	<i>individual contribution</i>		
4. Submission date:	<i>2024-05-23</i>		
5. Requester's reference (if applicable):			
6. Choose one of the following:			
This is a complete proposal:			<i>x</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>x</i>		
Name of the existing block:	<i>Bengali and Assamese</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	B.1-Specialized (small collection)	B.2-Specialized (large collection)	
C-Major extinct	D-Attested extinct	E-Minor extinct	
F-Archaic Hieroglyphic or Ideographic		G-Obscure or questionable usage symbols	
	<i>x</i>		
4. Is a repertoire including character names provided?			
<i>Yes</i>			
a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?			
<i>Yes</i>			
b. Are the character shapes attached in a legible form suitable for review?			
<i>Yes</i>			
5. Fonts related:			
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?			
<i>Jan Kučera if needed (glyph in Bengali block can be reused)</i>			
b. Identify the party granting a license for use of the font by the editors (include address, e-mail, ftp-site, etc.):			
6. References:			
a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?			
<i>Yes</i>			
b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?			
<i>Yes</i>			
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)?			

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: 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)

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:	Yes <i>scholars</i>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference:	No
4. The context of use for the proposed characters (type of use; common or rare) Reference:	<i>historic</i> <i>provided evidence</i>
5. Are the proposed characters in current use by the user community? If YES, where? Reference:	No
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:	Yes No
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	No
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:	Yes 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