

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation Internationale de Normalisation
 Международная организация по стандартизации

Doc Type: Working Group Document**Title: Results of the ad-hoc meeting on Assamese in London, 2018-06-19****Source: Michael Everson****Status: Ad-hoc report****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2018-06-21**

A discussion of the representation of the alphabet used to write Assamese was discussed at this meeting. The following points summarize the discussion.

- It is recognized that the Assamese language and the Bengali language pronounce many of the shared letters differently. It is also noted that the script is used for other languages: Bishnupriya Manipuri, Manipuri, Sylheti, Hajong, Rabha, and Deuri.
- It is recognized that UCS character names are immutable identifiers, and cannot be changed.
- It is recommended that Assamese character names be added to the names list as annotations, following this pattern:

0987 BENGALI LETTER I
 = Assamese letter hraswa i
 0988 BENGALI LETTER II
 = Assamese letter dirgha i

- It may be possible to change the block header name, though the block property values cannot. The most neutral and least disruptive name would be “Bengali-Assamese”. This is an editorial, not a normative, matter.
- A preliminary examination was made of the characters in the Assamese proposal which are not encoded was made in order to facilitate further work. Broadly speaking, the following notes were made. Further study is required. The Bureau of Indian Standards has offered to help coordinate this study.
 - Several recommendations for character additions (KHYA, HRASWA U (2nd form), DIRGHA U (2nd form), RI (2nd form), YA KAAR, REF, and RA KAAR) appear to be normal font presentation forms. KA-KAAR/NA-KAAR may or may not be, and requires further study.
 - Comparison of URDHA BINDU and NIMNA BINDU with the existing BENGALI ABBREVIATION SIGN needs to be done. One of these is likely unifiable with it.
 - The shape of ISSHAR (svargiya) may be different in Assamese (swargadeo).
 - There is an Assamese om which differs from Devanagari om.
 - The historical AVAGRAHA may differ in use and shape for Assamese (LUPTA A KAAR).
 - There are at least 11 characters used to represent land areas and fractions of land areas.
 - Additional independent vowels and vowel signs may be required. Assamese [ɔ] and [o] can be distinguished as in কলা *kôla* [kôla] (with the inherent vowel) ‘deaf’ and ক’লা *kola* [kola] ‘black’. Some additional sources (at Wiktionary) have shown a distinction also with between [ɛ] and [e] as in চক *sek* [sek] ‘Czech’ from চক’ক *sék* [sek] ‘cheque’ where

U+0027 APOSTROPHE or U+2019 RIGHT SINGLE QUOTATION MARK may be used. This may not be advantageous to Assamese. If atomic encoding were chosen for this, four characters might be required.

অ ô (exists) — (the inherent vowel)

অ' o (missing) ○' (missing)

ঐ ê (exists) ঠ○ (exists)

ঐ' e (missing) ঠ○' (missing)

It is worth noting, however, that MODIFIER LETTER APOSTROPHE has been used for similar purposes in Devanagari for Dogri and some other languages. A solution that suits both Devanagari and Bengali-Assamese scripts should be considered, whether that is using MODIFIER LETTER APOSTROPHE or using script-specific characters.

Participants in the ad-hoc:

Deborah Anderson

Chen Zhuang

Chui-Hsing Chiu

Peter Constable

Michael Everson

Karljürgen Feuerherm

Martin Hosken

Kim Gowoon

Prof. Kyongsok Kim

Toshiko Kimura

Jan Kučera

Alain LaBonté

Qin Lu

Manikandan K.

Karl Pentzlin

Prof. Shikhar Kumar Sarma

Michel Suignard

Toshiya Suzuki

Wataru Takagi

Shuichi Tashiro

Takada Tomokazu

Uma Umamaheswaran

Wei Lin-Mei

Yang Wangsung

Andrew West

Prof. Wu

M. K. Yadava





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

The Bengali script is also known as Bangla. In Assam, the name of the script is Asamiya or Assamese. The Assamese language has also been written historically using distinct regional scripts known as Kamrupi. The script is also used for Bishnupriya Manipuri, Meitei Manipuri, Sylheti, and Hajong.

Various signs






- 0980 ৭ BENGALI ANJI
= siddham, siddhirastu
• used at the beginning of texts as an invocation
- 0981 ৐ BENGALI SIGN CANDRABINDU
= Assamese sign sandrabindu
- 0982 ৑ BENGALI SIGN ANUSVARA
= Assamese sign anuswar
- 0983 ৒ BENGALI SIGN VISARGA
= Assamese sign bixarga

Independent vowels

- 0985 অ BENGALI LETTER A
= Assamese letter a
- 0986 আ BENGALI LETTER AA
= Assamese letter aa
- 0987 ই BENGALI LETTER I
= Assamese letter hraswa i
- 0988 ঐ BENGALI LETTER II
= Assamese letter dirgha i
- 0989 উ BENGALI LETTER U
= Assamese letter hraswa u
- 098A উ BENGALI LETTER UU
= Assamese letter dirgha u
- 098B ঋ BENGALI LETTER VOCALIC R
= Assamese letter ri
- 098C ৞ BENGALI LETTER VOCALIC L
- 098D  <reserved>
- 098E  <reserved>
- 098F এ BENGALI LETTER E
= Assamese letter e
- 0990 ঐ BENGALI LETTER AI
= Assamese letter oi
- 0991  <reserved>
- 0992  <reserved>
- 0993 ও BENGALI LETTER O
= Assamese letter o
- 0994 ঔ BENGALI LETTER AU
= Assamese letter ou

Consonants

- 0995 ক BENGALI LETTER KA
= Assamese letter ka
- 0996 খ BENGALI LETTER KHA
= Assamese letter kha
- 0997 গ BENGALI LETTER GA
= Assamese letter ga
- 0998 ঘ BENGALI LETTER GHA
= Assamese letter gha
- 0999 ঙ BENGALI LETTER NGA
= Assamese letter nga
- 099A চ BENGALI LETTER CA
= Assamese letter pratham sa
- 099B ছ BENGALI LETTER CHA
= Assamese letter dwitiya sa
- 099C জ BENGALI LETTER JA
= Assamese letter bargiya za
- 099D ঝ BENGALI LETTER JHA
= Assamese letter zha









- 099E ঞ BENGALI LETTER NYA
= Assamese letter nia
- 099F ট BENGALI LETTER TTA
= Assamese letter murdhanya ta
- 09A0 ঠ BENGALI LETTER TTHA
= Assamese letter murdhanya tha
- 09A1 ড BENGALI LETTER DDA
= Assamese letter murdhanya da
- 09A2 ঢ BENGALI LETTER DDHA
= Assamese letter murdhanya dha
- 09A3 ণ BENGALI LETTER NNA
= Assamese letter murdhanya na
- 09A4 ত BENGALI LETTER TA
= Assamese letter ta
- 09A5 থ BENGALI LETTER THA
= Assamese letter tha
- 09A6 দ BENGALI LETTER DA
= Assamese letter da
- 09A7 ধ BENGALI LETTER DHA
= Assamese letter dha
- 09A8 ন BENGALI LETTER NA
= Assamese letter na
- 09A9  <reserved>
- 09AA প BENGALI LETTER PA
= Assamese letter pa
- 09AB ফ BENGALI LETTER PHA
= Assamese letter pha
- 09AC ব BENGALI LETTER BA
= Bengali va, wa
= Assamese letter ba
- 09AD ভ BENGALI LETTER BHA
= Assamese letter bha
- 09AE ম BENGALI LETTER MA
= Assamese letter ma
- 09AF য BENGALI LETTER YA
= Assamese letter antahstha za
- 09B0 র BENGALI LETTER RA
- 09B1  <reserved>
- 09B2 ল BENGALI LETTER LA
= Assamese letter la
- 09B3  <reserved>
- 09B4  <reserved>
- 09B5  <reserved>
- 09B6 শ BENGALI LETTER SHA
= Assamese letter taalabya xa
- 09B7 ষ BENGALI LETTER SSA
= Assamese letter murdhanya xa
- 09B8 স BENGALI LETTER SA
= Assamese letter dantya xa
- 09B9 হ BENGALI LETTER HA
= Assamese letter ha

Various signs

- 09BC ৏ BENGALI SIGN NUKTA
• for extending the alphabet to new letters
- 09BD ৘ BENGALI SIGN AVAGRAHA







Dependent vowel signs

- 09BE ৑ BENGALI VOWEL SIGN AA
= Assamese vowel sign aa
- 09BF ি BENGALI VOWEL SIGN I
• stands to the left of the consonant
= Assamese vowel sign hraswa i
- 09C0 ী BENGALI VOWEL SIGN II
= Assamese vowel sign dirgha i

- 09C1  BENGALI VOWEL SIGN U
= Assamese vowel sign hraswa u
- 09C2  BENGALI VOWEL SIGN UU
= Assamese vowel sign dirgha u
- 09C3  BENGALI VOWEL SIGN VOCALIC R
= Assamese vowel sign ri
- 09C4  BENGALI VOWEL SIGN VOCALIC RR
= Assamese vowel sign dirgha ri
- 09C5  <reserved>
- 09C6  <reserved>
- 09C7  BENGALI VOWEL SIGN E
• stands to the left of the consonant
= Assamese vowel sign e
- 09C8  BENGALI VOWEL SIGN AI
• stands to the left of the consonant
= Assamese vowel sign oi

Two-part dependent vowel signs


These vowel signs have glyph pieces which stand on both sides of the consonant; they follow the consonant in logical order, and should be handled as a unit for most processing.

- 09CB  BENGALI VOWEL SIGN O
≡ 09C7  09BE 
= Assamese vowel sign o
- 09CC  BENGALI VOWEL SIGN AU
≡ 09C7  09D7 
= Assamese vowel sign ou

Virama

- 09CD  BENGALI SIGN VIRAMA
= hasant (Bengali term for halant)
= Assamese letter hasanta











Additional consonant

- 09CE  BENGALI LETTER KHANDA TA
• a dead consonant form of ta, without implicit vowel, used in some sequences
= Assamese letter khandata





Sign

- 09D7  BENGALI AU LENGTH MARK
= Assamese ou length mark

Additional consonants

- 09DC  BENGALI LETTER RRA
≡ 09A1  09BC 
= Assamese letter dare ra
- 09DD  BENGALI LETTER RHA
≡ 09A2  09BC 
= Assamese letter dhare ra
- 09DE  <reserved>
- 09DF  BENGALI LETTER YYA
≡ 09AF  09BC 
= Assamese letter antahstha ya

Additional vowels for Sanskrit











- 09E0  BENGALI LETTER VOCALIC RR
= Assamese letter dirgha ri
- 09E1  BENGALI LETTER VOCALIC LL
- 09E2  BENGALI VOWEL SIGN VOCALIC L
- 09E3  BENGALI VOWEL SIGN VOCALIC LL

Reserved



For viram punctuation, use the generic Indic 0964 and 0965. Note that these punctuation marks are referred to as dahri and double dahri in Bangla, and as daari, zati, and xiki, and as purnazati and aadhaa in Assamese.

- 09E4  <reserved>
→ 0964 I devanagari danda
- 09E5  <reserved>
→ 0965 II devanagari double danda



Digits

- 09E6  BENGALI DIGIT ZERO
= Assamese digit zero
- 09E7  BENGALI DIGIT ONE
= Assamese digit one
- 09E8  BENGALI DIGIT TWO
= Assamese digit two
- 09E9  BENGALI DIGIT THREE
= Assamese digit three
- 09EA  BENGALI DIGIT FOUR
= Assamese digit four
- 09EB  BENGALI DIGIT FIVE
= Assamese digit five
- 09EC  BENGALI DIGIT SIX
= Assamese digit six
- 09ED  BENGALI DIGIT SEVEN
= Assamese digit seven
- 09EE  BENGALI DIGIT EIGHT
= Assamese digit eight
- 09EF  BENGALI DIGIT NINE
= Assamese digit nine

Additions for Assamese







- 09F0  BENGALI LETTER RA WITH MIDDLE DIAGONAL
= Assamese letter ra
- 09F1  BENGALI LETTER RA WITH LOWER DIAGONAL
= Assamese letter wa
= bengali letter va with lower diagonal (1.0)

Currency signs


- 09F2  BENGALI RUPEE MARK
= taka
• historic currency sign
- 09F3  BENGALI RUPEE SIGN
= Bangladeshi taka

Historic symbols for fractional values

The use of these signs is not limited to currency, despite the character names.

- 09F4  BENGALI CURRENCY NUMERATOR ONE
• not in current usage
- 09F5  BENGALI CURRENCY NUMERATOR TWO
• not in current usage
- 09F6  BENGALI CURRENCY NUMERATOR THREE
• not in current usage
- 09F7  BENGALI CURRENCY NUMERATOR FOUR
- 09F8  BENGALI CURRENCY NUMERATOR ONE LESS THAN THE DENOMINATOR
- 09F9  BENGALI CURRENCY DENOMINATOR SIXTEEN

Sign

- 09FA  BENGALI ISSHAR
= ishvar
• represents the name of a deity
= svargiya
• written before the name of a deceased person

Historic currency sign

09FB ‹ BENGALI GANDA MARK

Signs

09FC ৎ BENGALI LETTER VEDIC ANUSVARA

09FD ° BENGALI ABBREVIATION SIGN