

**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 for Assigning a Distinct Code Point to Bangla YA-phalaa

2. Requester's name: Gautam Sengupta

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

4. Submission date: 2004-08-06

5. Requester's reference (if applicable): Attached document

6. Choose one of the following:
This is a complete proposal: Yes (taken with accompanying document)
or, More information will be provided later: More can be provided upon request.

B. Technical – General

1. Choose one of the following:
a. This proposal is for a new script (set of characters): No
Proposed name of script: N/A
b. The proposal is for addition of character(s) to an existing block: Yes
Name of the existing block: Bengali

2. Number of characters in proposal: 1

3. Proposed category (select one from below - see section 2.2 of P&P document):
A-Contemporary X 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 _____

4. Proposed Level of Implementation (1, 2 or 3) (see Annex K in P&P document): 1
Is a rationale provided for the choice? Yes
If Yes, reference: Not a combining character

5. 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

6. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard? Sayamindu Dasgupta (True Type)
If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used: <http://savannah.nongnu.org/download/freebangfont/Akaash-0.8.5.tar.gz> , sayamindu@clai.net

7. 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

8. 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 section D)

9. 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 <http://www.unicode.org/Public/UNIDATA/UCD.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.


¹ Form number: N2652-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11)

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before? If YES explain <u>N/A</u>	<u>No</u>
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? <u>Government of India, users and implementers in India and Bangladesh</u> If YES, available relevant documents: <u>http://tdil.mit.gov.in/prop_uni/new_bengali.pdf</u>	<u>Yes</u>
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included? Reference: <u>http://www.ethnologue.com/show_language.asp?code=BNG</u>	<u>Yes</u>
4. The context of use for the proposed characters (type of use; common or rare) Reference: <u>http://bangla.name/isis/download/ILCAA-2003.pdf</u>	<u>common</u>
5. Are the proposed characters in current use by the user community? If YES, where? Reference: <u>Newspapers, books, pamphlets</u>	<u>Yes</u>
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? <u>Yes</u> If YES, reference: <u>Addition to existing script block</u>	<u>Yes</u>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<u>N/A</u>
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? <u>N/A</u> If YES, reference: <u>N/A</u>	<u>No</u>
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? <u>Yes</u> If YES, reference: <u>Existing representation is inadequate. See section E.</u>	<u>Yes</u>
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character? If YES, is a rationale for its inclusion provided? <u>N/A</u> If YES, reference: <u>N/A</u>	<u>No</u>
11. Does the proposal include use of combining characters and/or use of composite sequences? If YES, is a rationale for such use provided? <u>N/A</u> If YES, reference: <u>N/A</u> Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided? <u>N/A</u> If YES, reference: <u>N/A</u>	<u>No</u>
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) <u>N/A</u>	<u>No</u>
13. Does the proposal contain any Ideographic compatibility character(s)? <u>No</u> If YES, is the equivalent corresponding unified ideographic character(s) identified? <u>N/A</u> If YES, reference: <u>N/A</u>	<u>No</u>

D. Proposed Characters

One character is proposed: general category and case mapping properties are as shown

Code position	Glyph	Name	Character properties
09FB		BENGALI LETTER YAPHALAA	<p>The general category for this character is Lo. In default collation, it should have the same weight as BENGALI LETTER YA. Other Unicode character properties match those of other similar characters, such as BENGALI LETTER YA.</p> <p>Behaves like the final consonant in an aksara (orthographic syllable) with respect to maatraa-placement; but does not ligate with a preceding consonant or take a below-base maatraa or reph. These must attach to the preceding consonant. A Yaphalaa can never be the first character in an aksara.</p>

E. Background Information

1 Introduction

In TUS 4.0 the Bangla YA-phalaa is assumed to be an allograph or display variant of the abstract character YA, and accordingly both are assigned to the same code point U+09AF. The purpose of this brief note is to argue that Ya-phalaa is an independent grapheme in Bangla and needs to be encoded as such with a code point distinct from that assigned to YA. The advantages accruing from adopting such a measure are discussed and highlighted.

2 Justification

Although Bangla YA and YA-phalaa are both diachronically related to the Brahmi YA, in the course of development of Bangla orthography the two graphs shifted so far apart that there seems to be no reason whatsoever to consider them to be allographs of the same grapheme or display variants of the same abstract character. Such graphemic splits are common phenomena in the history of scripts. The distinction between Bangla YA and YYA also emerged in the same way. The following arguments conclusively demonstrate the independent status of YA-phalaa as a distinct grapheme in the Bangla script:

2.1 Contrastive Distribution

YA and YA-phalaa are in contrastive distribution. Both may occur in the same environment and the occurrence of the one or the other cannot be predicted on the basis of the context of occurrence:

- i অ্যা A+YA-phalaa+AA-kaar
- ii অযা A+YA+AA-kaar

Evidently both YA and YA-phalaa can occur between an A and an AA-kaar ruling out the possibility of context-based predictability of occurrence.

The same conclusion may be reached following a completely different line of argumentation:

- A (i) and (ii) must have distinct encoded forms since they have distinct rendered shapes.
- B (i) and (ii) are identical except for the occurrence of YA in one of them in a position that is occupied by Ya-phalaa in the other.
- C From A and B it follows that YA and YA-phalaa must have distinct encodings.
- D But, if YA and YA-phalaa have distinct encodings then they cannot be display variants of the same abstract character.

2.2 Non-overlapping Denotation

The character YA always represents the phoneme /j/ as in (ii) above, and the character YA-phalaa never does so. The latter either represents a front mid-vowel, as in (i) and (iii) or gemination of the preceding consonant, as in (iv):

- iii ব্যক্তি /becti/ “person”
- iv সত্য /jotto/ “truth”

YA and YA-phalaa share no common denotation. At all levels of linguistic structure distributional variants of any basic element must share a common denotational core. Allomorphs of the same morpheme must have identical meaning or function, allophones of the same phoneme must share a common set of distinctive feature, etc. The fact that YA and YA-phalaa have no denotational overlap clearly indicates that they cannot be allographs of the same grapheme or display variants of the same abstract character.

3 Advantages of Encoding YA-phalaa as a Distinct Character

The following advantages accrue from encoding Ya-phalaa as a distinct abstract character:

- অ্য and এ্য can be encoded as <A/E U+xxxx> without inserting a VIRAMA after the vowel
- য় can be encoded as <RA U+xxxx> and য় as <RA VIRAMA YA> resolving the ambiguity of <RA VIRAMA YA> under the current specifications.

It is important to realize that these are major advantages that accrue from adopting a proposal that does not specifically target them but is independently motivated on theoretical grounds.

5 Disadvantages

Since how YA-phalaa should be encoded in the various environments in which it occurs, e.g. after vowels, after RA etc., is still being actively debated, it is unlikely that the proposed change will break any existing implementation or create problems with backwards compatibility. After carefully considering all aspects of the issue and consulting experts in the area, the Government of India has also arrived at the conclusion that it is necessary to assign a distinct code point to YA-phalaa. The changes in the Bangla code chart endorsed by them can be found in the following document: http://tdil.mit.gov.in/prop_uni/new_bengali.pdf. The only disadvantage of adopting the present proposal would perhaps be the cost of adding a new character, but it is decisively outweighed by the advantages.

F. References

Books and Articles

Chatterjee, Suniti Kumar. *The Origin and Development of the Bengali Language*. Calcutta University Press, Calcutta: 1926.

Chatterjee, Suniti Kumar. *Bhasa-Prakash Bangala Vyakaran*. Rupa, New Delhi: 1945.

Dasgupta, Probal and Gautam Sengupta. "The Bangla-Asamiya Scripts and its Representation in Unicode. In: Bhaskararao, Peri (ed.) *Indic Scripts: Past and Future*. ILCAA, Tokyo: 2003.

Masica, Colin P. *The Indo-Aryan Languages*. Cambridge University Press, Cambridge: 1991.

Pal, Palash Baran. *Dhvanimala Varnamala*. Papyrus, Kolkata: 2001.

Sen, Sukumar. *Bhasar Itibritta*. Ananda Publishers, Kolkata: 1994.

Dictionaries

Bandyopadhyaya, Haricharan. *Bangiya Sabdakos*. Sahitya Academy, New Delhi: 1966.

Basu, Rajsekhar. *Calantika*. M.C. Sarkar, Kolkata.

Chaudhuri, Jamil. *Bangla Academy Bangla Banan-Abhidhan*. Bangla Academy, Dhaka: 2001.

Das, Jnanendramohan. *Bangala Bhasar Abhidhan*. Sahitya Sangsad, Kolkata: 1979.

Datta, Bijitkumar et al (eds.) *Akademi Vidyarthi Bangla Abhidhan*. Pascimbanga Bangla Academi, Kolkata: 2000.

WWW Resources

Ethnologue: Web. http://www.ethnologue.com/show_language.asp?code=BNG. SIL International, Dallas: 2004.