**PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS
FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646 / UNICODE**

A. Administrative

1.	Title:	**Revised Proposal for Encoding Syloti Nagri Script in the BMP**
2.	Requesters' names:	Peter Constable (SIL International); James Lloyd-Williams and Sue Lloyd-Williams (Sylheti Translation And Research, London); Advocate Shamsul Islam Chowdhury (Chairman, Sylot Academy, Sylhet, Bangladesh); Professor Asaddar Ali (Vice Chairman, Sylot Academy, Sylhet, Bangladesh); Mohammed Sadique (First Secretary, Embassy of Bangladesh, Stockholm, Sweden); Matiar Rahman Chowdhury (Chairman, Sylot Academy (UK and Europe), London).
3.	Requester type:	Expert contribution
4.	Submission date:	2003-05-10
5.	Requester's reference (if applicable):	Prior L2/UTC documents L2/02-387, L2/02-388, L2/03-146r.
6a.	Completion:	This is a complete proposal
6b.	More information to be provided	See referenced document in item 5

B. Technical—General

1a. New script? Name?

Yes. Syloti Nagri.

1b. Addition of characters to existing block? Name?

No.

2. Number of characters

45

3. Proposed category

Category A

4. Proposed level of implementation and rationale

Level 3. Syloti Nagri script contains combining diacritics.

5a. Character names included in proposal?

Yes.

5b. Character names in accordance with guidelines?

Yes.

5c. Character shapes attached in a reviewable form?

Yes.

6a. Who will provide computerized font?

Sue Lloyd-Williams (Sylheti Translation And Research)

6b. Font currently available?

Yes.

6c. Font format?

TrueType

7a. Are references (to other character sets, dictionaries, descriptive texts, etc) provided?

Yes. See L2/02-388.

7b. Are published examples (such as samples from newspapers, magazines, or other sources) of use of proposed characters attached?

Yes. See L2/02-388.

8. Does the proposal address other aspects of character data processing?

Yes. See L2/02-388.

C. Technical—Justification

1. Has this proposal for addition of character(s) been submitted before?

This is a revision resulting from a request by UTC 93 (action item 93-A102) to reconsider specific issues and make specific revisions.

2. Contact with the user community?

Yes.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Sizeable communities in Sylhet region of Bangladesh, in Calcutta, in England and elsewhere.

3b. Reference

See L2/02-388.

4a. The context of use for the proposed characters (type of use; common or rare)?

Publishing in the script was known to have been done by several presses in Bangladesh and Calcutta as recently as the 1970s. More recently, no metal-type facilities have been available, but development of digital type has resulted in renewed use and growing interest.

4b.Reference

See L2/02-388.

5. Are the proposed characters in current use by the user community?

Yes.

5b. Where?

In London and Birmingham, and in Sylhet, possibly elsewhere.

5c. Reference

See L2/02-388.

6. After giving due considerations to the principles in N 1352 must the proposed characters be entirely in the BMP?

Yes.

6b. Rationale

Living script.

6c. Reference

See L2/02-388.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

Yes. One possible exception is SYLOTI NAGRI SIGN FUL (U+xx28 in the chart): a similar character can be found in Bengali poetic texts, and thus it may make sense to unify this character across scripts. If unification were desired, then this character could be added to the General Punctuation block, as it is a punctuation character.

8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

9a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Syloti Nagri has similarities with other North Indic scripts, but is distinct; it is no more like any other Indic scripts already in the UCS than are any of them to one another. Thus, for purposes of this proposal, these characters cannot be considered similar to existing characters.

10a. Does the proposal include use of combining characters and/or use of composite sequences (see clause 4.11 and 4.13 in ISO/IEC 10646-1)?

Yes.

10b. If YES, is a rationale for such use provided?

Yes.

10c. Reference

See L2/02-388.

10d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

Yes.

10e. Reference

See L2/02-388.

11. 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).

Yes. (See additional notes below and further discussion in L2/02-388 and L2/03-146r.)

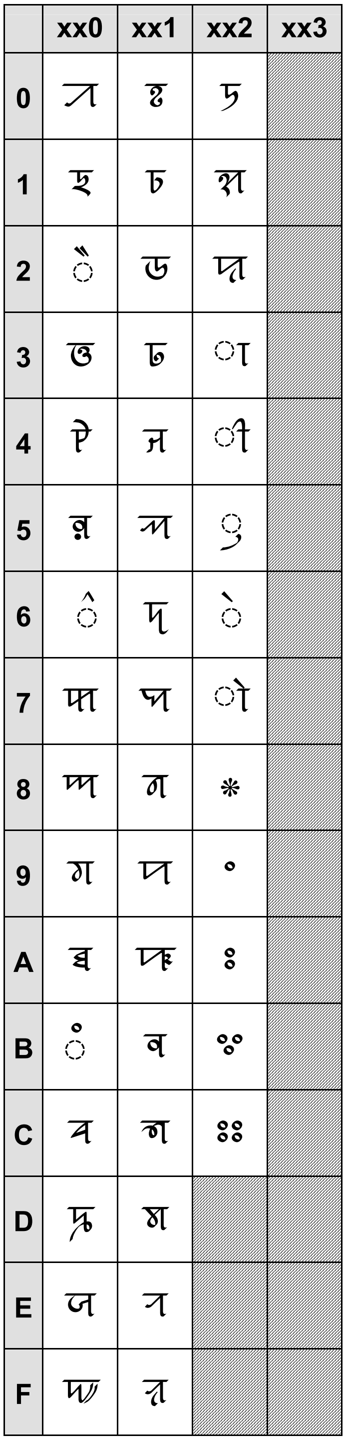
D. Proposed Characters

As discussed in section II.6 of L2/02-388, the basic collation order for Syloti Nagri is well established, but there are attested variations in ordering in relation to a few details. In the following chart, characters are ordered in what we believe to be the most preferable ordering.

As discussed in section II.1.6 of L2/02-388, there is very limited attestation in manuscripts of a set of Syloti Nagri digits. The evidence is not sufficient to support a proposal to encode Syloti Nagri digits at this time. It is conceivable that a proposal for Syloti Nagri digits may become justified at some future pont if further manuscripts are discovered, in which case an additional column of characters would be required. We have no reason to anticipate this happening in the near future, however.

(The code chart is presented on a new page.)

Code Chart:



Character Names:

U+xx00 SYLOTI NAGRI LETTER A
U+xx01 SYLOTI NAGRI LETTER I
U+xx02 SILOTI NAGRI SIGN DVISVARA
U+xx03 SYLOTI NAGRI LETTER U
U+xx04 SYLOTI NAGRI LETTER E
U+xx05 SYLOTI NAGRI LETTER O
U+xx06 SYLOTI NAGRI SIGN HASANTA
= halant, virama
U+xx07 SYLOTI NAGRI LETTER KO
U+xx08 SYLOTI NAGRI LETTER KHO
U+xx09 SYLOTI NAGRI LETTER GO
U+xx0A SYLOTI NAGRI LETTER GHO
U+xx0B SYLOTI NAGRI SIGN ANUSVARA
U+xx0C SYLOTI NAGRI LETTER CO
U+xx0D SYLOTI NAGRI LETTER CHO
U+xx0E SYLOTI NAGRI LETTER JO
U+xx0F SYLOTI NAGRI LETTERJHO
U+xx10 SYLOTI NAGRI LETTER TTO
U+xx11 SYLOTI NAGRI LETTER TTHO
U+xx12 SYLOTI NAGRI LETTER DDO
U+xx13 SYLOTI NAGRI LETTER DDHO
U+xx14 SYLOTI NAGRI LETTER TO
U+xx15 SYLOTI NAGRI LETTER THO
U+xx16 SYLOTI NAGRI LETTER DO
U+xx17 SYLOTI NAGRI LETTER DHO
U+xx18 SYLOTI NAGRI LETTER NO
U+xx19 SYLOTI NAGRI LETTER PO
U+xx1A SYLOTI NAGRI LETTER PHO
U+xx1B SYLOTI NAGRI LETTER BO
U+xx1C SYLOTI NAGRI LETTER BHO
U+xx1D SYLOTI NAGRI LETTER MO
U+xx1E SYLOTI NAGRI LETTER RO
U+xx1F SYLOTI NAGRI LETTER LO
U+xx20 SYLOTI NAGRI LETTER RRO
U+xx21 SYLOTI NAGRI LETTER SO
U+xx22 SYLOTI NAGRI LETTER HO
U+xx23 SYLOTI NAGRI VOWEL SIGN A
U+xx24 SYLOTI NAGRI VOWEL SIGN I
U+xx25 SYLOTI NAGRI VOWEL SIGN U
U+xx26 SYLOTI NAGRI VOWEL SIGN E
U+xx27 SYLOTI NAGRI VOWEL SIGN OO
U+xx28 SILOTI NAGRI SIGN FUL
U+xx29 SILOTI NAGRI POETRY MARK 1
U+xx2A SILOTI NAGRI POETRY MARK 2
U+xx2B SILOTI NAGRI POETRY MARK 3
U+xx2C SILOTI NAGRI POETRY MARK 4
U+xx2D..U+xx3F (These positions shall not be used.)

Unicode character properties

With the exception of the characters listed below, we propose that all characters have a General Category property of "Lo", a Canonical Combining Class property of 0, a Bidirectional Class property of "L", a Mirrored property of "N", an East Asian Width property of "N", and a Line Breaking property of "AL". In the information provided below, only those values that differ from the defaults just mentioned are listed.

Characters	Properties
U+xx06	General Category = "Mn", Canonical Combining Class = 9, Bidi Class = NSM, Line Breaking = "CM"
U+xx02, U+xx0B, U+xx26	General Category = "Mn", Canonical Combining Class = 230, Bidi Class = NSM, Line Breaking = "CM"
U+xx23, U+xx24, U+xx27	General Category = "Mc", Canonical Combining Class = 0, Line Breaking = "CM"
U+xx25	General Category = "Mn", Canonical Combining Class = 220, Bidi Class = NSM, Line Breaking = "CM"
U+xx28	General Category = "Po", Line Breaking = "QU"
U+xx29..U+xx2C	General Category = "Po", Line Breaking = "AL"
U+xx2D	General Category = "Cf", Line Breaking = "SA"

The canonical combining classes assigned to characters of general category Mc (other than hasanta / virama) do not match typical practice for Indic scripts (which would be 0). We propose these classes as they better deal with the behaviours of the script and are less likely to lead to problems in implementation and use. For instance, the DVISVARA and VOWEL SIGN U can co-occur, but these combining marks do not typographically interact, so alternate orderings of the characters cannot result in visible distinction and hence cannot have any distinct information value. By putting these two combining marks in distinct, non-zero combining classes, these two orderings become canonically equivalent.

Additional notes

One of the results of discussion of documents L2/02-387 and L2/02-388 at the UTC 93 meeting was that the proposed encoding model for dealing with conjunct formation was considered controversial. In L2/03-146r, various alternate encoding models for Syloti Nagri were discussed. Our proposal has been revised to use an encoding model like that used for Myanmar script (model "D" in L2/03-146r): U+xx06 SYLOTI NAGRI HASANTA is used to determine formation of a conjunct of the preceding and following characters, in which case the HASANTA has no direct visual display. If, however, the HASANTA is followed by ZWNJ, no conjunct is formed, and the HASANTA becomes visible.