

Request for encoding 11355 GRANTHA LENGTH MARK

Shriramana Sharma – jamadagni-at-gmail-dot-com

2009-Oct-25

This is a request for encoding a character in the Grantha block. While I have only recently submitted a proposal for the entire Grantha script, I chose to submit this proposal for this single character separately because of reasons I will state below.

The glyph ३ is seen in four different places in Grantha following independent and dependent vowels, its basic purpose being to lengthen the vowel. Ref 1 shows this usage of this glyph for lengthening the independent vowels U and Vocalic R and L and ref 2 does the same for the dependent vowel sign Vocalic L as seen below (left, ref 1, right, ref 2):

u	॒		क॒	क॒३
ū	॒३	& ॒३ & ॒३	ख॒	ख॒३
r	॒	॒	घ॒	घ॒३
ṛ	॒	॒३ & ॒३	॒	॒३
l	॒	॒	॒	॒३
ḷ	॒	॒३ & ॒३	॒	॒३

Sometimes it is also seen with a swash leading below the previous glyph, as in ref 3:

Grantha	॒	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३
Grantha	॒	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३	॒३

It should be noted that the independent vowels UU and vocalic LL (and the dependent vowel vocalic LL as well) have alternative representations which do not use this lengthening mark. (The alternative glyph for the dependent vowel for vocalic LL is the same as that of the independent vowel.)

Further its usage for lengthening the independent vowel U is archaic and not found in contemporary printings, which do show the other three usages however.

It might be useful to encode this glyph as a separate character to enable its independent depiction in texts that speak about the script. Such an encoding already has parallels in 0C55 TELUGU LENGTH MARK and 0CD5 KANNADA LENGTH MARK.

However, I still have some concerns. Both the Telugu and Kannada length marks are seen only in dependent vowel signs and not as parts of independent vowels. Even then, only the Kannada character is part of the decomposition of the precomposed long vowels of Kannada – 0CC0 KANNADA VOWEL SIGN II, 0CC7 KANNADA VOWEL SIGN EE and 0CCB KANNADA VOWEL SIGN OO. The Telugu equivalent, despite being quite a valid candidate to be used in the same way, has not been indicated as part of the decompositions of the corresponding Telugu vowel signs. In fact, these Telugu vowel signs have no decomposition at all!

Further, none of the independent vowels in Indic scripts have decompositions (with the sole exception of 0B94 TAMIL LETTER AU). In fact, the descriptions of all the other Indic scripts in TUS 5.0 chapter 9 contain recommendations *against* composing two-part independent vowels from their glyphic parts, such as composing 0906 DEVANAGARI LETTER AA as 0905 DEVANAGARI LETTER A + 093E DEVANAGARI VOWEL SIGN AA. This is possibly because these independent vowels were, for whatever reason, not provided decompositions which would ensure canonical equivalence. (The same is true for the Telugu long vowel signs as well.)

As for Grantha, only the independent vowel Vocalic RR consistently uses this length mark. The other three cases – independent vowel UU and independent and dependent vowels Vocalic LL – do not consistently show this length mark as they have alternative representations. Therefore it would not be appropriate even in Grantha to provide decompositions for those three cases with this proposed length mark as a component. Thus it is to be doubted whether the single remaining case of Vocalic RR should be decomposed.

Further, encoding this character will necessitate a recommendation against using the sequence short vowel + length mark to denote the long vowels. As the Sanskrit maxim goes: “prakṣālanāddhi paṅkasya dūrād asparśanaṃ varam” (It is better not to go near the mud at all rather than wash it off your clothes), I wonder whether it is worth enough to encode this glyph as a separate character. My technically-aware peers in the Grantha user community also were not very enthusiastic about encoding this separately. [These are the reasons why I did not include this character in the full proposal for Grantha.]

Despite this, I submit this request for encoding purely in consideration of the potential need to denote this glyph in plaintext and going by the precedent in Unicode of

what has been done for the Kannada and Telugu length marks (and the various AI and AU length marks in Bengali, Oriya etc).

Whether to encode as combining mark or not

In view of the previous discussion, this character is to be encoded purely for the need to show it independently in texts that speak about the script. Thus it is intended to stand in its own right and not in combination with any other character. Therefore I doubt that there is a need to encode it as a combining mark. Not encoding it with GC=Mc might also help prevent its usage as a modifier to vowels. As its original purpose is to denote language content (length of vowels) it is also not appropriate to give GC=So. Hence GC=Lo would be the logical choice. Thus the Unicode character properties for this character would be:

```
11355;GRANTHA LENGTH MARK;Lo;0;L; ; ; ; ;N; ; ; ;
```

If however there is an overriding need to encode it with GC=Mc, the UTC is free to do so. I also note that I have suggested the placement of this character at 11355 to be isomorphic with the corresponding Kannada and Telugu characters.

Acknowledgment

The idea of separately encoding this character came to me from Kent Karlsson's document L2/09-277 in which he suggests separately encoding this character. While I do not agree with many of his arguments (which chiefly are intended to unify Tamil and Grantha), I here encode this character for different reasons than his. Nevertheless I thank him for this idea.

References

1. South Indian Scripts in Sanskrit Manuscripts and Prints, Reinhold Grünendahl, Wiesbaden, Germany, 2001. ISBN 3-447-04504-3.
2. Samskrita Mutar Puttakam, Civappirakāca Paṇṭitar, Cōṭiṭappirakāca Yantira Cālai, Kokku, Śrī Laṅkā, 1991.
3. Ancient and Modern Alphabets of the Popular Hindu Alphabets of the Southern Peninsula of India, Capt Henry Harkness, 1837, Royal Asiatic Society, <http://www.archive.org/details/ancientmodernalp00harkrich>

Official Proposal Summary Form

A. Administrative

1. Title

Request for encoding 11355 GRANTHA LENGTH MARK

2. Requester's name

Shriramana Sharma

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

Individual contribution

4. Submission date

2009-Oct-25

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

Yes.

1d. Name of the existing block

Grantha

2. Number of characters in proposal

1 (one)

3. Proposed category:

Category A (Contemporary) or Category B1 (Specialized Small). See Grantha proposal by same author for details.

4a. Is a repertoire including character names provided?

Yes.

4b. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?

Yes.

4c. Are the character shapes attached in a legible form suitable for review?

Yes.

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Elmar Kniprath (kniprath-at-online-dot-de), Germany will provide the font for the Grantha proposal submitted by this author and that font contains the glyph required for this proposal as well.

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. 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.

8. 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.

See detailed proposal.

C. Technical – Justification

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

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes. Shriramana Sharma is part of the user community.

2b. If YES, with whom?

Vinodh Rajan, Chennai.

2c. If YES, available relevant documents

None specifically. Mode of contact was personal conversation.

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

Vedic scholars, Hindu temple priests and others interested in Sanskrit.

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

Quite commonly used for denoting some long vowels in the Grantha script.

4b. Reference

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

Yes.

5b. If YES, where?

In Tamil Nadu, Sri Lanka and elsewhere.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No. Because the Grantha script is being encoded in the SMP due to insufficient space in the BMP.

6b. If YES, is a rationale provided?

6c. If YES, reference

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

Only one character is proposed. It is to be placed isomorphically to the length marks of other blocks.

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

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

No.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

No.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No.

-O-O-O-