L2/10-341

Encoding sub-base Grantha vowel signs for Vocalic L and LL

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2010-Sep-14

§1. Background

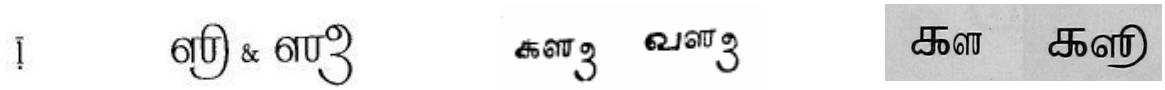
In my Grantha proposal L2/09-372 I had requested the distinct encoding of sub-base vowel signs for Vocalic L/LL at 1137E and 1137F while the *regular* vowel signs for Vocalic L/LL are to be encoded at 11362 and 11363 isomorphically with the other major Indic blocks.

The regular vowel signs are "regular" in that they are the written forms seen in most (contemporary) printings and writings. These regular forms are placed to the right of their base. However the very same glyphs as used for these regular forms are also attested to have been archaically used below their base:

VOWEL SIGN VOCALIC L		VOWEL SIGN VOCALIC LL	
REGULAR FORM	SUB-BASE FORM	REGULAR FORM	SUB-BASE FORM
◌𑌲	◌ 𑌲	◌𑌲𑌿	◌ 𑌲𑌿

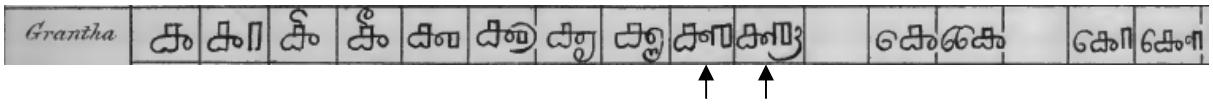
It is worthy noting here that in most other Indic scripts (starting from Brahmi, the progenitor of them all) the vowel signs for Vocalic L/LL are merely the same (glyph) as the independent vowel placed in sub-base position. This is true for all the three other major South Indian scripts supporting Sanskrit – Telugu, Kannada and Malayalam as well (of which Malayalam is furthermore a descendent of Grantha).

While the glyph shown in the Grantha code chart for the independent vowel Vocalic LL seems to be different from that used for the corresponding vowel sign as seen above, the written form chosen for the vowel sign in the code chart is but a ligated glyphic variant of the other, as shown (with proper references) in my proposal p 7:



Thus it is but natural that Grantha has these sub-base forms. In fact, it is only surprising that Grantha has any *other* forms. Anyhow, the fact remains that the "after-base" forms are

the regular forms seen in the majority of written Grantha. They are also the forms conducive to the occurrence of vowel sign ligatures as shown in my proposal p 20:



However, the sub-base forms have also been identified in printings, as in the following example provided in my proposal p 8:

Dr Gerhard Ehlers of the Orientabteilung of the Staatsbibliothek zu Berlin, who is an expert on manuscripts, confirmed to me in his email dated 2009-Sep-28 15:16 +0530 that "Vocalic L in Grantha is sometimes placed below, sometimes to the right" and he later confirmed in his mail dated 2009-Oct-19 16:17 +0530 that this is true for Vocalic LL as well. Thus while I was unable to obtain attestation from printings for the sub-base form of Vocalic LL, it is evident that such attestation is obtainable from manuscripts.

The scholars present at the 2010-Sep-06 meeting in the Indian Ministry of IT, New Delhi regarding the Grantha Unicode encoding also confirmed the existence of the sub-base variants of the vowel signs for Vocalic L/LL.

§2. The argument for encoding

The existence of two such forms of writing the vowel sign for Vocalic L/LL is somewhat unprecedented in Indic, and one must decide what to do about it in Unicode Grantha – whether to encode the sub-base forms as separate characters or not.

There is obviously no semantic difference between the two forms, as these are just different ways of denoting the same language/phonetic content. However, the variation here is not glyphic, nor is it that between ligated/unligated forms. To be precise, the variation is *positional*, and what's more it affects the potential GC of the forms in question. The sub-base forms, if encoded separately, would take GC=Mn as for the vowel signs for Vocalic L/LL in most other Indic scripts. The regular forms would obviously take GC=Mc.

Thus in the interest of providing a representation of the sub-base form in encoding, it would be appropriate that separate characters are encoded.

I note that in my proposal I had asked that canonical decompositions of these forms to the regular forms be provided to maintain semantic equivalence. However, only a character which looks identical to another character or group of characters is decomposed

this way, and “looks identical” does not apply to the present case due to the difference in glyph position. It is possible to maintain semantic equivalence for text search, collation etc using higher protocols as seen in the case of the BALINESE SURANG. Therefore decomposition should not be done and higher protocols must be used for this purpose.

Quite apart from the SURANG, even in Indic we have 094E DEVANAGARI VOWEL SIGN PRISHTHAMATRA E, which is semantically identical to 0947 DEVANAGARI VOWEL SIGN E. If at all a text would mixedly use these two vowel signs (for example a modern Devanagari text translating or explaining an archaic text using the prishthamatra) a programmer of higher protocols can always provide for their semantic identity. The present case of the two forms of the vowel signs may also be treated that way.

The only argument against separate encoding that I can think of is that it is better to not have two different encoded representations of the same phonetic content, but that argument falls against the precedent of the prishthamatra at 094E, which incidentally also has a different GC=Mc than that of its semantic equivalent at 0947 viz GC=Mn.

Malayalam and Grantha also already have two different encoded representations of the phonetic content /au/. The Devanagari prishthamatra, itself an alternative encoding of /e/, produces in combination with other characters alternative representations of the sounds /o/, /ai/ and /au/. In this situation, fault can hardly be found with creating alternative encoded representations for vocalic L and LL if sufficient justification is found.

§3. Conclusion

I am here only trying to present arguments which may logically and un-biasedly be put forth for the distinct encoding of these sub-base vowel signs. Since the presence of these written forms has been attested, and since it might be useful and technically justified (based on the GC) to provide for such written forms, I am obliged to say that at least from my viewpoint the evidence and arguments seem to weigh in favour of distinct encoding.

If the UTC so decides in favour of distinct encoding, the suggested codepoints, suggested names and Unicode character properties are:

```
1137E;GRANTHA VOWEL SIGN SUB-BASE VOCALIC L;Mn;0;NSM;;;;;N;;;;;
1137F;GRANTHA VOWEL SIGN SUB-BASE VOCALIC LL;Mn;0;NSM;;;;;N;;;;;
```

§4. Official Proposal Summary Form

A. Administrative

1. Title

Encoding sub-base Grantha vowel signs for Vocalic L and LL

2. Requester's name

Shriramana Sharma

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

Individual contribution

4. Submission date

2010-Sep-14

5. Requester's reference (if applicable)

6. Choose one of the following: This is a complete proposal (or) More information will be provided later

This is a complete proposal.

B. Technical – General

1. Choose one of the following:

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

No.

1b. The proposal is for addition of character(s) to an existing block, Name of the existing block

Yes. Grantha.

2. Number of characters in proposal

2 (two)

3. Proposed category

Category A contemporary (though sometimes considered archaic).

4. Is a repertoire including character names provided?

Yes.

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

Yes.

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

Yes.

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

Elmar Kniprath.

b. Identify the party granting a license for use of the font by the editors (include address, e-mail etc.)

Elmar Kniprath. kniprath - online - de (mail address obfuscated on request).

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.

These characters were part of the characters proposed for the Grantha script in L2/09-372. However, it became necessary to submit a separate proposal because these two characters did not form part of the consensus between the different Grantha proposals.

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. The proposer himself is also a member of the user community.

2b. If YES, with whom?

Ministry of IT, Govt of India. Sanskrit scholars of Tamil Nadu mentioned in L2/09-372 pp 56-57. Dr Gerhard Ehlers, Orientabteilung, Staatsbiliothek zu Berlin, Germany.

2c. If YES, available relevant documents

None specifically. The matter was discussed personally/via email.

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

See L2/09-372.

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

Rare.

4b. Reference

See proposal.

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

Yes albeit rarely.

5b. If YES, where?

Tamil Nadu, India. To a certain extent in Sri Lanka.

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

No.

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)?

Yes, since it is only logical to keep mutually related characters together.

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?

Yes. The proposed characters are glyphically similar to the "regular" vowel signs for the same Vocalic L/LL proposed by L2/09-372 at 11362 and 11363.

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

Yes.

10c. If YES, reference

See detailed proposal.

11a. Does the proposal include use of combining characters and/or use of composite sequences?

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)?

If YES, is the equivalent corresponding unified ideographic character(s) identified?

If YES, reference:

No.

-o-o-o-