

## Request for editorial updates to various Indic scripts

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

2012-Mar-17

### §1. Generic Indic

The phonological sequence /r vocalic\_r/ occurs now and then in Sanskrit. While one would expect this sequence to be written in the normal way as <consonant vowel\_sign>:

रु

... it is customary to instead write it as the independent\_vocalic\_r with a reph on top:

रृ

... as seen in various Sanskrit texts: (thanks: Mihail Bayaryn, <http://svayambhava.org>)

। á, ि i, ी î, -उ u, ू ū, -ए e, ै ai,  
। o, ी au, z. B.: का ká, कि ki, की kî, कु ku, कू kû, कृ kṛi,  
कृ kṛî, के ke, कै kai, को ko, कौ kau; रु ru, रू rû, रृ r-ri;  
शु oder सु śu, शू oder सू śû, शृ śṛi, हु oder ऊ hu, हृ hṛi.

कृत्यैर्ऋणे ॥ ४३ ॥

Mit einem Particip. fut. pass., wenn von Schulden die Rede ist.

मासदेयमणम्.

This is true not only of Devanagari but also of other Indic scripts. See for example the following sample from a Kannada printing of a Vedic (Sanskrit) text:

ಅಗ್ನಿಃ ಪೂರ್ವೇಭಿರ್ಯುಷ್ಭಿರೀಡ್ಯೋ ನೂತನೈರುತ |  
ಸ ದೇವಾ ಏಹ ವಕ್ಷತಿ ||೨||

In principle this is extended to other “vocalic” vowels also:

ऋ ऌ ॠ ॡ

The opinion has been heard from Sanskrit scholars that the encoded sequence LETTER RA + VOWEL SIGN VOCALIC R/RR/L/LL should be used to encode these written forms as the underlying phonemic content is such. However, Unicode encodes written forms and not the sound values thereof. The observed written form is that of a reph attached to an independent vowel (in Devanagari and other scripts). In all Indic scripts that exhibit it, the reph is represented by the encoded sequence LETTER RA + VIRAMA. As such, it is found advisable from the technical point of view to represent the written forms by the sequence:

LETTER RA + VIRAMA + LETTER VOCALIC R/RR/L/LL

The sequences LETTER RA + VOWEL SIGN VOCALIC R/RR/L/LL should not be used to represent the above written forms and their nominal expected display is:

ॠ ॡ ॢ ॣ

**Action Requested:**

As generic Indic encoding principles are set out in the Devanagari chapter, it is requested to document therein that when the reph is used with independent vowels to represent the phonological sequences /r vocalic\_r/ /r vocalic\_rr/ etc in traditional Sanskrit orthography in any Indic script, so:

ऋ ऌ ॠ ॡ

... the sequences:

LETTER RA + VIRAMA + LETTER VOCALIC R/RR/L/LL

... should be used for representing such written forms.

## §2. Tamil

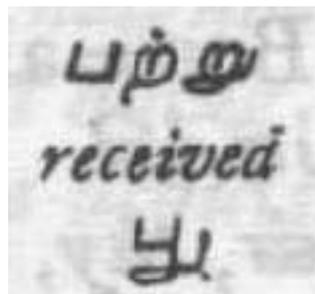
This is about the glyphs of two symbols: 0BF6 TAMIL DEBIT SIGN and 0BF7 TAMIL CREDIT SIGN.

0BF6    யு    TAMIL DEBIT SIGN  
          = patru

The TAMIL DEBIT SIGN is an abbreviation of the Tamil word *parru* and hence derived from the glyph of its first written letter *pa* viz 0BAA TAMIL LETTER PA. Thus the body of the glyph for this character should be corrected to correspond to the glyph of 0BAA:

0BAA    ட    TAMIL LETTER PA

As it is, the glyph of 0BF6 is much too narrow. Please observe the original written form from G U Pope's "A Handbook of the Tamil Language" (original 1855, 5th AES Reprint p 14; thanks: G Balachandran of Sri Lanka):



It is requested to modify the glyph accordingly to keep in with the origin of this character.

As regards 0BF7 TAMIL CREDIT SIGN, the current glyph looks like a ligature of 0B8E TAMIL LETTER E ஃ and 0BB5 TAMIL LETTER VA வ:

0BF7    ஃவ    TAMIL CREDIT SIGN  
          = varavu

However the glyph originally proposed in L2/01-375R N2381R looks like a ligature of 0B8E TAMIL LETTER E ஃ and 0BF3 TAMIL DAY SIGN ஁.

ஃ஁    |    TAMIL CREDIT SIGN (Varavu)

... as per the attestation provided in that document.

Apparently due to an inadvertent error the current glyph has been devised for the charts. This has also been reflected in the TN Govt's document on Unicode (L2/10-318, pp 15, 27). However it is certainly inaccurate in face of the attestation and should be corrected.

**Action Requested:**

It is requested to correct the two glyphs to match the actual attestation as described above.

### §3. Malayalam

This is about the Malayalam Dot Reph. TUS 6.0 p 310 (p 342 of PDF) says:

*U+0D4E MALAYALAM LETTER DOT REPH is used to represent the dead consonant form of U+0D30 MALAYALAM LETTER RA, when it is displayed as a dot over the consonant following it. Conceptually, the dot reph is analogous to the sequence <RA, VIRAMA>, but when followed by another consonant, the Malayalam cluster <RA, VIRAMA, C2> normally assumes the C2 conjoining form. U+0D4E MALAYALAM LETTER DOT REPH occurs first, in logical order, even though it displays as a dot above the succeeding consonant. It has the character properties of a letter, and is not considered a combining mark.*

There are two points to be corrected here. First, the original Dot Reph proposal L2/09-245R N3676 only says (on p 3) that "... the Malayalam cluster <RA, VIRAMA, C2> can assume C2-conjoining form." It is incorrect to state that the sequence <RA, VIRAMA, C2> normally assumes the C2 form.

In fact, experimenting with all possible <RA, VIRAMA, C2> sequences of Malayalam with commonly seen system fonts like Lohit Malayalam, Rachana (both Linux) and Kartika (Windows XP) shows that only for  $C2 \in \{YA, RA, LA, VA\}$  does the above sequence produce a C2-conjoining form and for all other consonants it merely produces [ra virama c2]. Therefore this is not the proper justification for encoding a dot reph\*.

---

\* To elaborate on this, looking back at the DOT REPH proposal, I find that it merely states that RA + VIRAMA could not be used to represent the reph because currently RA + VIRAMA + C2 is sometimes rendered with C2 assuming a conjoining form.

However even in Telugu the fact is that in the contemporary orthography RA + VIRAMA + C2 is rendered with C2 taking a conjoining form and in the old orthography there was a reph. See L2/12-017. The difference between the two orthographies in Telugu is recommended to be taken care of by altering the fonts.

It is hence recommended to remove this statement from the documentation. The rest of the text then deals with the placement of the dot reph.

Even here, the wording is insufficient. The fact is that this character should be treated like a real reph of other Indic scripts and reordered appropriately. (Vide Andrew Glass’s mail to Unicore dated Wed, 7 Mar 2012 16:36:33 +0000.) To be precise, it should be reordered before any post-base conjoining consonants or vowel signs, and placed centered on top of the consonant ligature/stack that acts as the base of the orthographic syllable:

ഉർദ്ധ്വോ	ഭരത്ര്യൗ
“ūrdhvyō”	“bhartryau”

Further in passing it is noted that the “dot” reph is sometimes also presented (as seen in the above rendering) as a small vertical stroke, due to which it is called the “gopi” reph colloquially. (“gopi” = a Hindu religious mark in sandal paste worn on one’s forehead.) It would be useful to document this stylistic variant as well.

**Action Requested:**

It is hence requested to change the passage quoted above as follows:

*U+0D4E MALAYALAM LETTER DOT REPH is used to represent the dead consonant form of U+0D30 MALAYALAM LETTER RA, when it is displayed as a dot or small vertical stroke above the consonant following it. Conceptually, the DOT REPH character U+0D4E is analogous to the sequence <RA, VIRAMA> which is rendered as a reph in most other Indic scripts. It is hence reordered in rendering just as the reph glyph of other Indic scripts would be, even though it occurs first in logical order. It has the character properties of a letter, and is not considered a combining mark.*

---

Even in Malayalam the fact is that in the older orthography the reph was used and in the contemporary orthography it is not used. However the difference between Malayalam and Telugu is that in Telugu it is possible to recommend the sequence RA + VIRAMA + ZWJ to select a reph even in contemporary style fonts whereas in Malayalam that sequence has considerable legacy use to represent the chillu form of R/RR and hence cannot be reassigned to denote the reph.

In addition the following passage on p 311 (p 343 of PDF):

*Words that formerly used dot reph are now spelled using U+0D7C Malayalam Chillu RR or the respective C2-conjoining forms.*

should also be corrected as follows due to reasons outlined above:

*Words that formerly used dot reph are now spelled using U+0D7C MALAYALAM CHILLU RR or simply the forms produced by RA + VIRAMA.*

#### §4. Oriya

In the Oriya script the sacred syllable OM is written as the letter O with a candrabindu. The candrabindu may either stand alone on top of the letter O, or it may ligate with the latter:



UNJOINED FORM



JOINED FORM 1



JOINED FORM 2

The evidence for these forms has been documented in L2/11-258.

Initially a separate character was requested to be encoded for the ligated form (L2/11-258) but based on later feedback from the user community (L2/11-343) it is now recommended that the ligated form be treated as a full equivalent of the unligated form, and the font be free to present the sequence O + CANDRABINDU as ligated or not.

In a font that would normally present O + CANDRABINDU as unligated, a ZWJ may be included to request a ligated form, so: O + ZWJ + CANDRABINDU. Likewise, in a font that would normally present O + CANDRABINDU as *ligated*, a ZWNJ may be included to request an *unligated* form, so: O + ZWNJ + CANDRABINDU.

#### **Action Requested:**

It is requested to document the existence of the ligated form of O + CANDRABINDU as the Oriya OM, alongwith the above model of selecting either form (if the model is found appropriate). The fact that the above model is based on the precedent of Bengali vowel sign ligatures (see TUS 6.0 pp 286,287; pp 318,319 of PDF) must be remembered here.

## §5. Brahmi

This is a request to correct the glyph for 11034 BRAHMI LETTER LLA. The current glyph (which is as proposed in L2/07-342 and L2/08-277R) looks like 11033 BRAHMI LETTER HA:

11033    𑀧    BRAHMI LETTER HA  
11034    𑀧    BRAHMI LETTER LLA

However, this is nowhere attested. The Brahmi proposal itself states (on p 8) that:

*both the Tamil Brahmi and the Bhattiprolu 𑀧 are ... derived from the regular letter 𑀧, the former by adding a hook to the lower right of 𑀧, the latter by mirroring 𑀧 horizontally (while the north-Indian 𑀧 is derived from the letter 𑀧).*

Now for the chart it is the North Indian Brahmi forms that are used in cases where the South Indian ones (Tamil and Bhattiprolu) are different, and the actual attested form for LLA in North Indian Brahmi is (from IndoSkrit, <http://userpage.fu-berlin.de/falk/>):



... which is indeed derived from the North Indian Brahmi form of DDA:



Note that the above form of DDA is a later development from the DDA shown in the charts:

1101F    𑀧    BRAHMI LETTER DDA

... but LLA itself was a later development in Brahmi (as indicated by Salomon speaking about “𑀧a ... not represented in standard (Northern) Brahmi” in the centuries BCE\*).

Thus the glyphic form which is earliest attested for LLA in northern Brahmi as shown above does not correspond to the DDA which is present in the code charts. Despite this, it is the earliest form *attested* and hence it is the form that should be shown as the representative glyph even though it is anachronistic.

---

\* Indian Epigraphy, Richard Salomon, Oxford University Press, New York, 1998, p 36 ll 1,2

While I should concede that Indoskript provides some earlier samples for LLA from the years BCE which look like the old DDA with a bulge or circular formation below:



... the interpretation of this as LLA against DDA itself is not universally accepted, given Salomon's remark of the LLA not existing in North Indian Brahmi before the advent of CE and IndoSkript also noting under "Alternate Readings" that other scholars have read it as DDA. Further, by myself examining the estampages provided by IndoSkript for the above examples and the readings thereof from Prinsep\*, I was not able to convince myself that this was a true early form of LLA.

In the absence of evidence to that the currently shown glyph or the above doubtful pre-BCE glyph for LLA is indeed the Brahmi representation of LLA, and in the face of the evidence that the other form, to wit:



... is indeed the oldest uncontested attested form of LLA in North Indian Brahmi, I write:

**Action Requested:**

It is requested to correct the glyph for 11034 BRAHMI LETTER LLA to match its earliest attestation:



-o-o-o-

---

\* <http://www.archive.org/download/lesinscriptionsd02asokuoft/lesinscriptionsd02asokuoft.pdf>