

Changes to Core Specification for Indic scripts for Unicode 11.0

Srinidhi A
srinidhi.pinkpetals24@gmail.com



Sridatta A
sridatta.jamadagni@gmail.com

December 8, 2017

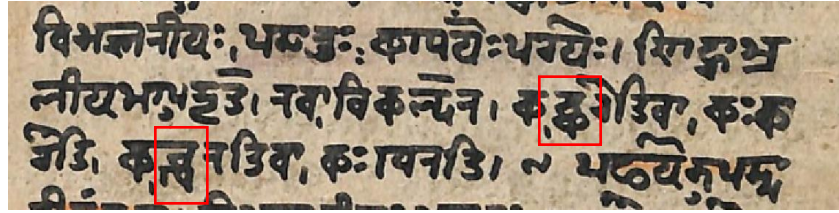
This document provides changes to Core Specification for Indic scripts. It was discussed in L2/17-098. Based on the recommendations of Script Ad Hoc, a separate document is presented here.

1 Sharada

In Sharada section of Core Specification, the following representation for Jihvamuliya is presented.

U+111C2  jihvamuliya + U+11191 क ka → कै
U+111C2  jihvamuliya + U+11192 ण kha → णै

However Jihvamuliya is not written above the consonant, instead it makes ligature with the consonant as कै. A sample of their representations is shown in below manuscript.



Jihvamuliya is various manuscripts. (from IndoSkript).

L2/17-255 asked “Does the change of shape for KA take place because of the jihvamuliya or because KA is subjoined below another consonant?”

In consonant clusters these letters behave similar to other consonants. The ligature occurs because KA is subjoined below another consonant.

Action Requested Correct the shape of combination of Jihvamuliya and Ka/Kha in core specification.

2 Kannada

In Kannada section of Core Specification, the following representation for Jihvamuliya is presented.

U+0CF1 ೀ jihvamuliya + U+0C95 ಕ ka → ೀ
U+0CF2 ಋ upadhmaniya + U+0CAB ಫ pha → ಫಠ

Dependent vowels signs can also be added to the stack:

U+0CF1 ೀ jihvamuliya + U+0C95 ಕ ka + U+0CBF ಿ vowel sign i → ೀಠ

Action Requested: ಹka should not have Talekattu (or headstroke). Correct the shape of ಹka from ೀ to ೀಠ (see L2/13-242 for attestations).

3 Sora Sompeng

Sora Sompeng is currently categorized as Abugida by Unicode. The following details are provided in page 614 of Core Specification.

“Encoding Structure. The Sora Sompeng script is structured as an abugida. The consonant letters contain an inherent vowel. There are no conjunct characters for consonant clusters, and there is no visible vowel killer to show the deletion of the inherent vowel. The reader must determine the presence or absence of the inherent schwa based on recognition of each word. The character repertoire does not match the phonemic repertoire of Sora very well.”

The proposal L2/09-189R says “Sora Sompeng is of the Brahmic type: the consonant letters contain an inherent vowel.”, but doesn’t provide attestations where the consonant letters contain inherent vowel as well as the actual usage of the script. The only attestation provided is an alphabet chart where the consonants are transcribed as s, t, b etc. without inherent vowel.

However after examining primary sources like books published by native speakers and name plates, it was concluded that Sora Sompeng is structurally an alphabet and cannot be categorized as abugida. It does not contain inherent vowel, virama, consonant conjuncts and vowel signs. Letters are represented using combination of consonants and vowels. For example Ka is represented as ಕಹ (KAH+AH) and Hargl (meaning ornament in Sora) is written as ಹಾಹರಾಹಲಾಹ (HAH+AH+RAH+GAH+LAH).



(https://upload.wikimedia.org/wikipedia/commons/c/cd/Mangei_Gomango.JPG)

Nameplate of Mangei Gomango who devised Sora Sompeng script at Odisha Tribal Development Society Bhubaneswar. Here Guru Mangei Gomango is written as ସୁସୁ ସିନିପିନି ସିନିନି.

“ ୠ ” (SA)

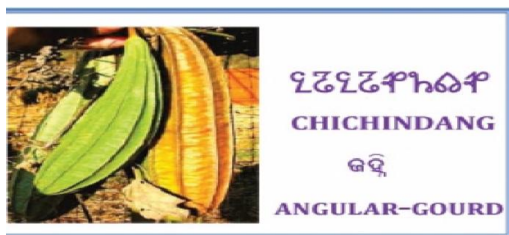
ୠ + ଇ = ୠଇ , ୠ + କ = ୠକ , ୠ + ଟ = ୠଟ
 S + A = SA , S + E = SE , S + I = SI
 ୠ + U = ୠU , ୠ + ୨ = ୠ୨
 S + U = SU , S + O = SO

୩ (TA)

୩ + ଇ = ୩ଇ , ୩ + କ = ୩କ , ୩ + ଟ = ୩ଟ
 T + A = TA , T + E = TE , T + I = TI
 ୩ + U = ୩U , ୩ + ୨ = ୩୨ .
 T + U = TU , T + O = TO .

A book called Savara Laglgn Abirnaba depicted above shows Combination of Consonants with vowels. Words in Sora Sompeng from the same book.¹

ୠୠୠୠ	SIING	୩୩୩୩୩୩	TANGNIY
ୠ୩୩ୠୠୠ	SANGSANG	୩୩୩୩୩୩	TANGNIY
ୠ୩ୠ୩୩	SEJAB	୩୩୩୩୩	TREDA
ୠୠୠୠ	SIING	୩୩୩୩୩୩୩	TIRIYUNG
ୠ୩୩୩୩	SURGA	୩୩୩୩୩	TUDUM
ୠ୩୩୩୩୩	SONETI	୩୩୩୩	TOGI



Words in Sora sompeng from a learning book written by Krushna Sabar.²

¹ <http://savara.bharatavani.in/book/savara-laglgn-abirnaba/>

Action Requested: Change the kind of writing system from abugida to alphabetic in Chapters 6.1, 15.14 and elsewhere. Update the section Encoding Structure of chapter 15.14 appropriately.

4 VEDIC SIGN NIHSHVASA

Section 4.5 of L2/09-372 and L2/09-298 informs U+1CD3 separates sections between which a pause is disallowed. The annotation in the code chart has been corrected as suggested in L2/09-372, but explanation in section 12.1 pg. 465 of the core specification “U+1CD3 VEDIC SIGN NIHSHVASA indicates where a breath may be taken.” is misleading. This should be corrected to “Separates sections of Sama Vedic singing between which a pause is disallowed” as in the code chart.

5 Ligature Forms for Ra + Vocalic Liquids

The Core specification gives following information of Ligature Forms for Ra + Vocalic Liquids. However these sequences are not restricted Devanagari, in fact it is attested in most Indic scripts which have repha. It is customary to instead write it as the independent_vocalic_r with a repha as noted in L2/12-106. Update the Core Specification indicating that these graphical forms are also attested in other Indic scripts.

For the benefit of font developers these formations may be mentioned in individual chapters or in this Devanagari section as ‘also attested in other Indic scripts’.

Ligature Forms for Ra + Vocalic Liquids. The phonological sequence /r vocalic_r/, expressed with the character sequence <U+0930 ra, U+0943 vocalic_r>, can graphically appear as either of two forms, as shown in the first row of *Table 12-4*. It may appear as the full independent vowel form of the vocalic_r, with a superscript *repha* form of the ra (V + RA_{sup}): रै. Alternatively, it may appear as the full letter form of the ra with the subscript, dependent form of the vocalic_r (RA_n + V_{vs}): रृ. Similarly, the phonological sequences with the other vocalic sounds (*rr, l, ll*) have two written forms, as shown in *Table 12-4*.

Table 12-4. RA + Vocalic Letter Ligature Forms

र	+	ॠ	→	रै	or	रृ
र	+	ॡ	→	रै	or	रृ
र	+	ॢ	→	रै	or	रृ
र	+	ॣ	→	रै	or	रृ

²<http://savara.bharatavani.in/book/savara-language-learning-book/>

ಇಂದ್ರೋ ವಹ್ನಿಃ ಪಿತೃಪತಿರ್ನೈರ್ಯುತೋ ವರುಣೋ ಮರುತ್ ಕುಬೇರ ಈಶಃ ಪತಯಃ ಪೂರ್ವಾದೀನಾಂ ದಿಶಾಂ ಕ್ರಮಾತ್	೯೨
ಐರಾವತಃ ಪುಂಡರೀಕೋ ವಾಮನಃ ಕುಮುದೋಽಂಜನಃ ಪುಷ್ಪದಂತಸ್ಸಾರ್ವಭೌಮಸ್ಸುಪ್ರತೀಕಶ್ಚ ದಿಗ್ಗಜಾಃ	೯೩
ಕರಿಣ್ಯೋಽಭ್ಯಮುಕಪಿಲಾಪಿಂಗಲಾನುಪಮಾಃ ಕ್ರಮಾತ್ ತಾಮ್ರಕರ್ಣೇ ಶುಭ್ರದಂತೀ ಚಾಂಗನಾ ಚಾಂಬನಾವತೀ	೯೪
ಕ್ಷೀಬಾವ್ಯಯಂ ತ್ವಪದಿಶಃ ದಿಶೋರ್ಮಧ್ಯೇ ವಿದಿಕ್ ಸ್ತ್ರಿಯಾಮ್ ಅಭ್ಯಂತರಂ ತ್ವಂತರಾಲಂ ಚಕ್ರವಾಲಂ ತು ಮಂಡಲಮ್	೯೫
ಅಭ್ಯಂ ಮೇಘೋ ವಾರಿವಾಹಃ ಸ್ತನಯಿತ್ನುರ್ಬಲಾಹಕಃ ಗಾರಾದರೋ ಜಲದರಪ ಟಿತಾನಾರಿದೋಽಂಬುಭೃತ್	೯೬

೨. ಅಗ್ನಿಃ ಪೂರ್ವೇಭಿರ್ಯುಷಿಭಿಃ—ರೀಡ್ಯೋ ನೂತನೈರುತ |

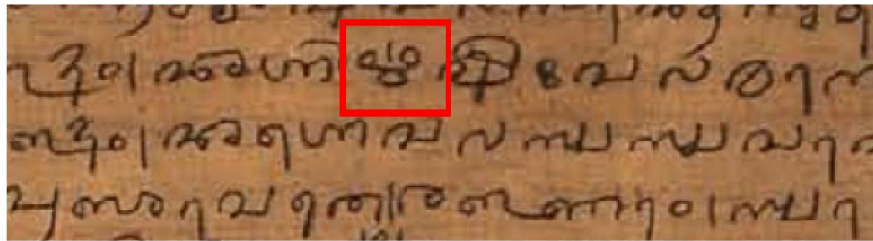
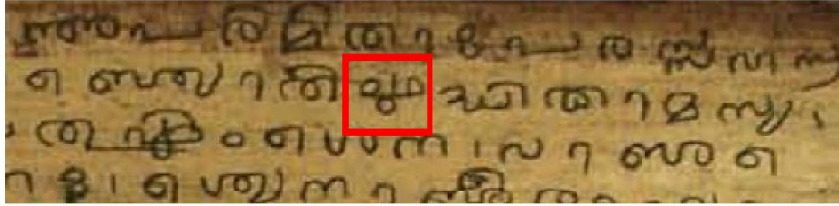
ಸ ದೇವಾ ಏಹ ವಕ್ತೃತಿ

|| ೨ ||

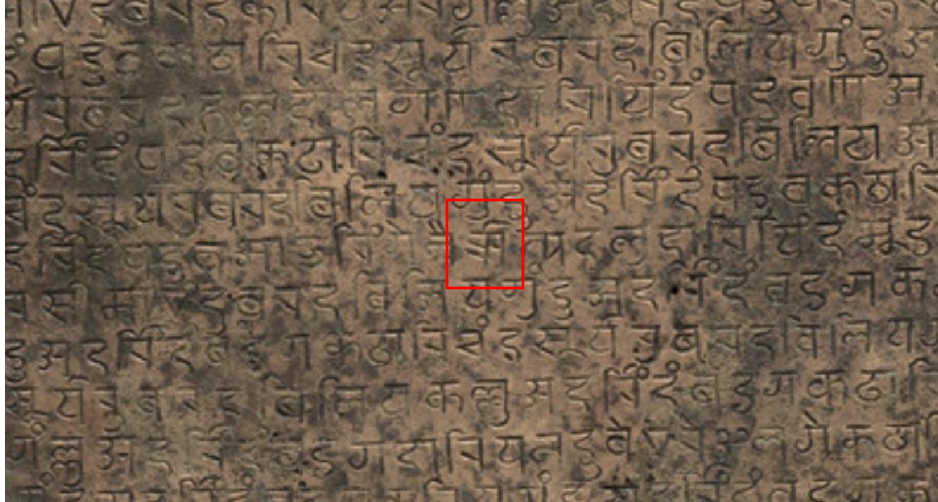
ಪುರಾಣ ಋಷಿಗಳು ಯಾವಾತನ ಸ್ತುತಿಯನ್ನು ಮಾಡಿದ್ದರೋ, ಈಗಿನ ಋಷಿಗಳು ಯಾವಾತನನ್ನು ಸ್ತುತಿಸುವರೋ, ಅಂಥ ಅಗ್ನಿಯು ದೇವತೆಗಳನ್ನು ಈ ಯಜ್ಞದಲ್ಲಿ ಆಹ್ವಾನಿಸಲಿ.

|| ೨ ||


rr in Kannada.






rr in Tigalari.









rr in Nandinagari.

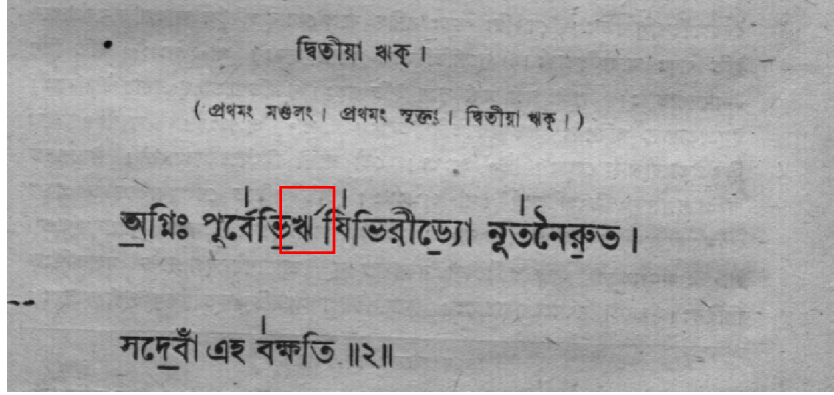
When  VOWEL SIGN VOCALIC R occurs with RA, the latter is written as *repha* and placed above the arrow-head of the vowel letter:

 rr < RA,  VOWEL SIGN VOCALIC R>

rr in Bhaiksuki (from L2/14-091).

	mva	म	म
	yva	य	य
	yva	य	य
	rr	र	र
	rka	क	क
	rkha	ख	ख

rr in Odia.



ṛṛ in Bengali.

nirṛṇa	নির্কৃষ্ণ	నిర్కృష్ణ	నిర్కృష్ణ
nirṛtiḥ	নির্কৃষ্টি:	నిర్కృష్టి:	నిర్కృష్టి:

ṛṛ in Telugu. Note that modern orthography does not use repha. It is attested in older texts.

Similar forms may also exist in other scripts which have repha like Tirhuta, Newa, Siddham, Gujarati etc.