

South Asia Subcommittee Report

February 5, 2008

The South Asia subcommittee regrets very much the passing away of Prof. Joshi and expresses its condolences to his family.

1 Dandas

The subcommittee recommends the following guidelines:

For the currently encoded scripts:

- if the orthographies using the script do not make use of dandas, don't encode dandas (e.g. Sinhala)
- if the script already has encoded dandas, those should be used in the context of that script (e.g. Devanagari)
- otherwise, a decision to encode script specific dandas or to use the dandas encoded for another script should be made and, to ensure the perenity of data, that choice should not be reversed. The use of dandas is generally enough to justify their encoding in the script.
- for Bengali, Gurmukhi, Gujarati, Oriya, Tamil, Telugu, Kannada, Syloti Nagri, for which DEVANAGARI dandas are currently used, the decision should be made at the earliest possible time, taking into account the existence of data

For new scripts:

- if the orthographies using the script do not make use of dandas, don't encode dandas
- otherwise, a decision to encode script specific dandas or to use the dandas encoded for another script should be made and, to ensure the perenity of data, that choice should not be reversed. The use of dandas is generally enough to justify their encoding in the script.

2 Malayalam

Recommendation:

In Unicode 5.1, change the current Table 2, "NTA Cluster":

		4b	$\text{ᦺ} + \text{ᦺ} + \underline{\text{R}}$ (oD7B, oD31, <u>R</u>)	hen <u>ry</u>
5	L ᦺ ᦺ ᦺ <u>R</u>	5	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \text{L} + \text{ᦺ} + \underline{\text{R}^2}$ (oD28, oD4D, oD31, L, oD31, <u>R</u> ²)	ce <u>menteri</u>
6	ᦺ L ᦺ ᦺ <u>R</u>	6a	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{LR}}$ (oD7B, oD31, oD4D, oD31, <u>LR</u>)	van <u>tony</u>
		6b	$\text{ᦺ} + \text{ᦺ} + \text{L} + \text{ᦺ} + \underline{\text{R}^2}$ (oD7B, oD31, L, oD31, <u>R</u> ²)	van <u>rera</u>
7	ᦺ ᦺ L ᦺ <u>R</u>	7a	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{LR}}$ (oD28, oD4D, oD31, oD31, <u>LR</u>)	me <u>ntarodu</u>
		7b	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{LR}}$ (oD7B, oD31, oD31, <u>LR</u>)	Mr. Be <u>nraro</u>
8	ᦺ ᦺ ᦺ <u>R</u>	8a	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{R}}$ (oD28, oD4D, oD31, <u>R</u>)	co <u>mmentary</u>
		8b	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{R}}$ (oD7B, oD31, oD4D, oD31, <u>R</u>)	Ame <u>rican Time</u>
		8c	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{R}}$ (oD7B, oD31, oD31, <u>R</u>)	Mr. Be <u>nrari</u>

to:

	Visual		Character Sequence	Reading
1	L ᦺ <u>R</u>	1a	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{LR}}$ (oD7B, oD4D, oD31, <u>LR</u>)	e <u>nte</u>
		1b	$\text{ᦺ} + \text{ZWJ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{LR}}$ (oD7B, 200D, oD4D, oD31, <u>LR</u>)	
	L ᦺ ᦺ <u>R</u>	1c	$\text{ᦺ} + \text{ᦺ} + \text{ZWJ} + \text{ᦺ} + \underline{\text{LR}}$ (oD7B, oD4D, 200D, oD31, <u>LR</u>)	
3	ᦺ L ᦺ <u>R</u>	3	$\text{ᦺ} + \text{ᦺ} + \underline{\text{LR}}$ (oD7B, oD31, <u>LR</u>)	e <u>nroll</u>
4	ᦺ ᦺ <u>R</u>	4a	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \underline{\text{R}}$ (oD37, oD4D, oD31, <u>R</u>)	e <u>ntamme</u>
		4b	$\text{ᦺ} + \text{ᦺ} + \underline{\text{R}}$ (oD7B, oD31, <u>R</u>)	he <u>nry</u>
5	L ᦺ ᦺ ᦺ <u>R</u>	5	$\text{ᦺ} + \text{ᦺ} + \text{ᦺ} + \text{L} + \text{ᦺ} + \underline{\text{R}^2}$ (oD28, oD4D, oD31, L, oD31, <u>R</u> ²)	ce <u>menteri</u>

6		6a	 $(oD7B, oD31, oD4D, oD31, LR)$	van <u>tony</u>
		6b	 $(oD7B, oD31, L, oD31, R^2)$	van <u>rera</u>
7		7a	 $(oD28, oD4D, oD31, oD31, LR)$	men <u>tarodu</u>
		7b	 $(oD7B, oD31, oD31, LR)$	Mr. Ben <u>raro</u>
8		8a	 $(oD28, oD4D, oD31, R)$	com <u>mentary</u>
		8b	 $(oD7B, oD31, oD4D, oD31, R)$	Ameri <u>c</u> an Time
		8c	 $(oD7B, oD31, oD31, R)$	Mr. Ben <u>rari</u>
9		9	 $(oD7B, oD31, oD31, R)$	

Change Table 4a, “Dot Reph RA+ YA” from the current:

	Visual	Proposed Encoding in 5.1	Alternative Encoding
1		CHILLU RR, YA (oD7C, oD2F)	same
2		RA, VIRAMA, YA (oD30, oD4D, oD2F)	same
3		RA, VIRAMA, ZWNJ, YA (oD30, oD4D, 200C, oD2F)	same
4		<i>unrepresentable</i>	CHILLU RR, VIRAMA, YA (oD7C, oD4D, oD2F)
5		RA, VIRAMA, YA, VIRAMA, YA (oD30, oD4D, oD2F, oD4D, oD2F)	CHILLU RR, VIRAMA, YA, VIRAMA, YA (oD7C, oD4D, oD2F, oD4D, oD2F)

to:

	Visual	Character sequence
1		CHILLU RR, YA

2	ॠ	RA VIRAMA YA RA ZWJ VIRAMA YA
3	ॡ	RA VIRAMA ZWNJ YA
4	ॢ	RRA VIRAMA ZWJ YA
5	ॣ	RRA VIRAMA YA RRA ZWJ VIRAMA YA
6	।	RRA VIRAMA SA RRA VIRAMA ZWJ SA
6	॥	RRA VIRAMA YA VIRAMA YA

3 Vedic

The subcommittee considered the latest iterations of the proposals, specifically L2/o8-050 from Everson/Scharf and L2/o8-042 as amended by L2/o8-097 from the Government of India, and was pleased to see a great level of convergence.

In general, the subcommittee followed the code points and names proposed in L2/o8-050, and included the names proposed in L2/o8-042 as annotations. To facilitate the work of the UTC, many of the recommendations are formulated as resolutions and action items that the UTC could adopt. The recommended action items of the form “update L2/o8-050 ...” would result in the updated L2/o8-050 that matches the recommended resolutions. In a few cases, the committee did not arrive at a complete consensus, but identified a likely outcome or majority opinion.

The subcommittee believes that after adoption of its recommendations, the encoding of Vedic is ready for balloting.

3.1 Prishthamatra

In the Prishthamatra orthography, a number of vowel sounds are written in dependent form using different shapes than in the modern orthography:

sound	Prishthamatra	modern
ke	ॠक	के
kai	ॡके	के
ko	ॢका	को
kau	ॣको	कौ

Three approaches have been considered:

- 1a. encoding of a combining character (ॠ) for the left side part, and use of the existing vowel signs for the other parts

1b. encoding of four combining characters (0970-3) for the four vowel signs, to be consistent with the current practice of “atomic” encoding of the vowel signs

2. use the same sequences as for the modern representation and a higher protocol to trigger a rendering in one or the other orthography.

sound	Solution 1	Solution 1a	Solution 1b	Solution 2
ke	क	<0915, 09x0>	<0915, 09y0>	<0915, 0947>
kai	के	<0915, 09x0, 0947>	<0915, 09y1>	<0915, 0948>
ko	का	<0915, 09x0, 093E>	<0915, 09y2>	<0915, 094B>
kau	को	<0915, 09x0, 094B>	<0915, 09y3>	<0915, 094C>

The subcommittee believes that solution 1a would be preferable to solution 1b. It is also noted while solution 2 is technically feasible, it would not allow the distinction of orthographies in plain text, and would require significant extensions to the existing rendering implementations.

Majority opinion:

The UTC resolves to:

- accept U+094E DEVANAGARI VOWEL SIGN PRISHTHAMATRA E

3.2 Rigvedic

Recommendation:

The UTC resolves to:

- accept U+1CE0 VEDIC TONE RIGVEDIC KASHMIRI INDEPENDENT SVARITA with the annotation *vaidika uurdhva vakra rekhaa*

Action item:

- update L2/08-050 to:
 - add the annotation above

3.3 Samavedic, letters and digits

Majority opinion:

The UTC resolves to:

- accept U+A8E0 COMBINING DEVANAGARI DIGIT ZERO with the annotation *vaidika saamasvara anka shuunya*
- accept U+A8E1 COMBINING DEVANAGARI DIGIT ONE with the annotation *vaidika saamasvara anka eka udatta*
- accept U+A8E2 COMBINING DEVANAGARI DIGIT TWO with the annotation *vaidika saamasvara anka dvi svarita*
- accept U+A8E3 COMBINING DEVANAGARI DIGIT THREE with the annotation *vaidika saamasvara anka tri anudatta*
- accept U+A8E4 COMBINING DEVANAGARI DIGIT FOUR with the annotation *vaidika*

- saamasvara anka chatur*
- accept U+A8E5 COMBINING DEVANAGARI DIGIT FIVE with the annotation *vaidika saamasvara anka pancha*
- accept U+A8E6 COMBINING DEVANAGARI DIGIT SIX with the annotation *vaidika saamasvara anka shatt*
- accept U+A8E7 COMBINING DEVANAGARI DIGIT SEVEN with the annotation *vaidika saamasvara anka sapta*
- accept U+A8E8 COMBINING DEVANAGARI DIGIT EIGHT with the annotation *vaidika saamasvara anka ashta*
- accept U+A8E9 COMBINING DEVANAGARI DIGIT NINE with the annotation *vaidika saamasvara anka nava*
- accept U+A8EA COMBINING DEVANAGARI LETTER A with the annotation *vaidika saamasvara abhinihita*
- accept U+A8EB COMBINING DEVANAGARI LETTER U with the annotation *vaidika saamasvara u*
- accept U+A8EC COMBINING DEVANAGARI LETTER KA with the annotation *vaidika saamasvara ka*
- accept U+A8ED COMBINING DEVANAGARI LETTER NA with the annotation *vaidika saamasvara namana*
- accept U+A8EE COMBINING DEVANAGARI LETTER PA with the annotation *vaidika saamasvara prannatam*
- accept U+A8EF COMBINING DEVANAGARI LETTER RA with the annotation *vaidika saamasvara ra*
- accept U+A8F0 COMBINING DEVANAGARI LETTER VI with the annotation *vaidika saamasvara vinata*
- accept U+A8F1 COMBINING DEVANAGARI SIGN AVAGRAHA with the annotation *vaidika saamasvara dirghibhava*

Action item:

- update L2/08-050 to:
 - add the annotations above

3.4 Samavedic, others

Majority opinion:

The UTC resolves to:

- accept U+1CD0 VEDIC TONE KARSHANA with the annotation *vaidika saamasvara karshanna*
- accept U+1CD1 VEDIC TONE SHARA with the annotation *vaidika svarita uurdhva shara*
- accept U+1CD2 VEDIC SIGN PRENKHA with the annotation *vaidika saamasvara prenkha*
- accept U+1CD3 VEDIC SIGN NIHSHVASA with the annotation *vaidika saamagaana yogakaala*, as a non-combining character, with the general category Po

Action items:

- update L2/08-050 to
 - update the annotation of U+1CD3
 - remove the proposed U+1CD4 VEDIC SIGN KAMPA
- investigate the appropriate linebreak property assignment to prevent a linebreak before U+1CD3

3.5 Yajurvedic, general

Recommendation:

The UTC resolves to:

- add the annotation *vaidika svarita adhah konna* to U+1CD6 VEDIC TONE YAJURVEDIC INDEPENDENT SVARITA
- move U+1CD8 VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA to U+1CD7, and to add the annotation *vaidika svarita adho vakra rekhaa*
- move U+1CD7 VEDIC TONE YAJURVEDIC AGGRAVATED INDEPENDENT SVARITA to U+1CD5, and to add the annotation *vaidika svarita adho nyubja*
- move U+1CDA VEDIC TONE CANDRA BELOW to U+1CD8, and to add the annotation *vaidika svarita adhah ardha vakra*
- move U+1CD4 VEDIC TONE DOUBLE SVARITA to U+1CDA, and to add the annotation *vaidika svarita uurdhva dvi rekhaa*
- move U+1CD5 VEDIC TONE TRIPLE SVARITA to U+1CDB, and to add the annotation *vaidika svarita uurdhva tri rekhaa*
- move U+1CDF VEDIC TONE DOT BELOW to U+1CDD, and to add the annotation *vaidika svarita adho bindu*
- move U+1CE0 VEDIC TONE KATHAKA ANUDATTA to U+1CDC, and to add the annotation *vaidika svarita adho rekhaa*
- accept U+1CD9 VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA SCHROEDER, with the annotation *vaidika svarita adho samyukta rekhaa*

The net result is:

- U+1CD5 VEDIC TONE YAJURVEDIC AGGRAVATED INDEPENDENT SVARITA = *vaidika svarita adho nyubja*
- U+1CD6 VEDIC TONE YAJURVEDIC INDEPENDENT SVARITA = *vaidika svarita adhah konna*
- U+1CD7 VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA = *vaidika svarita adho vakra rekhaa*
- U+1CD8 VEDIC TONE CANDRA BELOW = *vaidika svarita adhah ardha vakra*
- U+1CD9 VEDIC TONE YAJURVEDIC KATHAKA INDEPENDENT SVARITA SCHROEDER = *vaidika svarita adho samyukta rekhaa*
- U+1CDA VEDIC TONE DOUBLE SVARITA = *vaidika svarita uurdhva dvi rekhaa*
- U+1CDB VEDIC TONE TRIPLE SVARITA = *vaidika svarita uurdhva tri rekhaa*
- U+1CDC VEDIC TONE KATHAKA ANUDATTA = *vaidika svarita adho rekhaa*
- U+1CDD VEDIC TONE DOT BELOW = *vaidika svarita adho bindu*

Action item:

- update L2/08-050 to
 - update the annotation of U+1CD6, 1CD7, 1CD9, 1CDA, 1CDB as above

3.6 Yajurvedic, Satapathabrahmana

Recommendation:

The UTC resolves to:

- move U+1CDD VEDIC TONE THREE DOTS BELOW to U+1CDF, and to add the annotation *vaidika svarita adhas tri bindu*
- add the annotation *vaidika svarita adho dvi bindu* to U+1CDE VEDIC TONE TWO DOTS BELOW

The net result is:

- U+1CDE VEDIC TONE TWO DOTS BELOW = *vaidika svarita adho dvi bindu*
- U+1CDF VEDIC TONE THREE DOTS BELOW = *vaidika svarita adhas tri bindu*

Action item:

- update L2/08-050 to

- update the annotation of U+1CDF as above

Recommendation:

Action items:

- discuss further the proposed character U+1CEF VAIDIKA KAMPA from L2/08-042
- investigate further the proposed character U+1CF0 VAIDIKA SVARITA URASI REKHAA from L2/08-042, to obtain more evidence about the use and form of this character

3.7 Atharvavedic

Recommendation:

The UTC resolves to:

- move U+1CDB VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA to U+1CE1, and to add the annotation *vaidika svarita dvi vakra khanda*

The net result is:

- U+1CE1 VEDIC TONE ATHARVAVEDIC INDEPENDENT SVARITA = *vaidika svarita dvi vakra khanda*

3.8 Ardhavisarga

Recommendation:

The UTC resolves to:

- accept U+1CE2 VEDIC SIGN VISARGA SVARITA with the annotation *vaidika madhyerekhaa*
- accept U+1CE3 VEDIC SIGN VISARGA UDATTA with the annotation *vaidika dakshinnatah uurdhvaga*
- accept U+1CE4 VEDIC SIGN REVERSED VISARGA UDATTA with the annotation *vaidika vaamatah uurdhvaga*
- accept U+1CE5 VEDIC SIGN VISARGA ANUDATTA with the annotation *vaidika vaamatah adhoga*
- accept U+1CE6 VEDIC SIGN REVERSED VISARGA ANUDATTA with the annotation *vaidika dakshinnatah adhoga*
- accept U+1CE7 VEDIC SIGN VISARGA UDATTA WITH TAIL with the annotation *vaidika dakshinnatah uurdhva vakra*
- accept U+1CE8 VEDIC SIGN VISARGA ANUDATTA WITH TAIL with the annotation *vaidika vaamatah adho vakra*

Action item:

- update L2/08-050 to
 - use the annotation above for U+1CE2

Recommendation:

The UTC resolves to:

- accept U+1CF1 VEDIC SIGN ARDHAVISARGA with the annotation *vaidika jihvaamuliya upadhamaaniya*, as a spacing combining character

Action item:

- update L2/o8-050 to
 - make U+1CF1 a combining character
 - add the annotation above to U+1CF1

3.9 Nasals

Recommendation:

The UTC resolves to:

- accept U+0900 DEVANAGARI SIGN INVERTED CANDRABINDU with the annotation *vaidika adhomukha candrabindu* and the Script property Devanagari (same as the existing U+0901 DEVANAGARI SIGN CANDRABINDU)
- accept U+A8F2 DEVANAGARI SIGN SPACING CANDRABINDU with the annotation *vaidika candrabindu*
- add the annotation *vaidika anusvaara candrabindu tiryak* to U+A8F3 DEVANAGARI SIGN CANDRABINDU VIRAMA
- add the annotation *vaidika anusvaara dvi candrabindu tiryak* to U+A8F4 DEVANAGARI SIGN DOUBLE CANDRABINDU VIRAMA
- add the annotation *vaidika anusvaara candrabindu sadvi* to U+A8F5 DEVANAGARI SIGN CANDRABINDU TWO
- add the annotation *vaidika anusvaara candrabindu satri* to U+A8F6 DEVANAGARI SIGN CANDRABINDU THREE
- add the annotation *vaidika anusvaara candrabindu saavagraha* to U+A8F7 DEVANAGARI SIGN CANDRABINDU AVAGRAHA

Action item:

- update L2/o8-050 to:
 - move the proposed U+A8F2 at U+0900
 - move the proposed U+A8F3..A8F5 up one slot at U+A8F2..A8F4
 - remove the proposed U+A8F6
 - move the proposed U+A8F7..A8F9 up two slots at U+A8F5..A8F
 - add the annotations above

Further investigation: because L2/o7-343 shows examples of vamagomukha with tail without a bindu or candrabindu, it is likely that the Everson/Scharf approach is preferable. We would want to hear more from the Govt. of India, but there is a priori a majority opinion:

The UTC resolves to:

- accept U+1CE9 VEDIC SIGN ANUSVARA ANTARGOMUKHA with the annotation *vaidika anusvaara antarmukha*
- accept U+1CEA VEDIC SIGN ANUSVARA BAHIRGOMUKHA with the annotation *vaidika anusvaara naagaphannaa*
- accept U+1CEB VEDIC SIGN ANUSVARA VAMAGOMUKHA with the annotation *vaidika anusvaara vaamamukha*
- accept U+1CEC VEDIC SIGN ANUSVARA VAMAGOMUKHA WITH TAIL with the annotation *vaidika anusvaara vaamamukha savakra*

Action item:

- update L2/o8-050 to:
 - change the name of U+1CEC to the name above (remove the “R” in VAMAGOMUKHA)

Recommendation:

Action item:

- further discuss the proposed character U+1CF7 VAIDIKA ANUSVAARA UBHAYATO MUKHA of L2/o8-04

Recommendation:

The UTC resolves to:

- accept U+1CEE VEDIC SIGN HEXIFORM LONG ANUSVARA with the annotation *vaidika anusvaara anugami*
- accept U+1CEF VEDIC SIGN LONG ANUSVARA with the annotation *vaidika anusvaara dakshinnamukha*
- accept U+1CF0 VEDIC SIGN RTHANG LONG ANUSVARA with the annotation *vaidika anusvaara tthasadrisha*

Action item:

- update L2/o8-050 to:
 - change the annotations as above

3.10 Additions for Devanagari

Recommendation:

The UTC resolves to:

- accept U+0973 DEVANAGARI SIGN PUSHPIKA with the annotation *vaidika pushpika*
- accept U+0974 DEVANAGARI CARET with the annotation *vaidika trutikaa*, as non-combining character

Action item:

- update L2/o8-050 to:
 - change the annotations as above

Recommendation: There seems to be no problem in representing the written form of the phoneme *llha* by <LLA, VIRAMA, HA>. This seems to be just the equivalent of a digraph (but in the context of an Indic Script, hence the *virama*).

3.11 Miscellaneous

Recommendation:

The UTC resolves to:

- accept U+1CED VEDIC SIGN TIRYAK with the annotation *vaidika tiryak*

Action item:

- update L2/o8-050 to:
 - change the annotations as above

3.12 Jaiminiya

Recommendation: Jaiminiya cantillation can be delimited using the existing interlinear annotation characters U+FFFB - U+FFFC (see section 16.8 of the standard, which discusses the use of those characters; also note that those characters are appropriate for use with any script).