South Asia Subcommittee

Encoding of Vedic

1 Introduction

The purpose of this page is to record the current state of proposals, discussions, and consensus on the encoding of Vedic. The discussions take place on the southasia@unicode.org mailing list.

2 Documents

The Everson, Scharf et al. proposals:

- 2008-01-28: [L2/08-050](#)
- 2007-10-18: [L2/07-343](#); differences with L2/07-230 in [L2/07-394](#); additional evidence in [L2/08-035](#)
- 2007-07-26: [L2/07-230](#); history of that document in [L2/07-262](#)
- 2007-04-26: [L2/07-095R](#)
  - Comments in [L2/07-388](#)
    - Comments in [L2/07-400](#)

The Government of India proposals:

- 2008-01-28: [L2/08-042](#), [L2/08-043](#)
- 2007-10-10: [L2/07-395](#), [L2/07-396](#), [L2/07-397](#)
- 2006-05-10 [L2/06-185](#)

Comparasions:

- 2007-10-18: [L2/07-401](): L2/07-343 vs. L2/07-39{5,6,7}
- 2007-08-08: [L2/07-271](): L2/07-230 vs. L2/06-185
  - Response in [L2/07-386]()
    - Comments in [L2/07-400]()

3 Areas of work

The reminder of this page lists the various areas that need work. It follows the organization of L2/07-343, so as to able to work on manageable chunks.

Each area starts with the relevant characters proposed in L2/07-343, L2/08-050, L2/07-39x and L2/08-042. Those that have been accepted during the October 2007 UTC meeting #113 are preceded with a green "y".

This is followed by a summary of the discussion on the list. When a discussion item has been resolved by latest iterations of the proposal, that discussion has been omitted. In other words, only the discussion relevant to the current disagreements is included.

3.1 Prishthamatra

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

sound	Prishthamatra	modern
ke		
kai		
ko		
kau		

Three approaches have been considered:

1. encoding of a combining character for the left side part, and use of the existing vowel signs U+0947 DEVANAGARI VOWEL SIGN E, U+093E DEVANAGARI VOWEL SIGN AA and U+094B DEVANAGARI VOWEL SIGN O.
2. encoding of four combining characters for the four vowel signs
3. use of the same sequences as for the modern representation and and higher-protocol to trigger a rendering in one or the other orthography.

L2/08-050 (and earlier L2/07-230) and Peter Constable favor 1:

accepted	Everson/Scharf		Govt. India		name(s)
	07-343	08-050	07-39x	08-042	
	T◌	94e			devanagari vowel sign prishthamatra e

The Government of India favors 3. Sobi, on the southasia list, 2008-01-03:

> [...] we have already shown serious concern about the code point in our comments document L2/07-388, observation No 3. We still retain our same stands that Prishthamatra is not advisable to be separately encoded.

> In reference to the comment made in 2.1 in document no L2/07-343 we are of the opinion that font may handle such situation however it may need rendering engine reordering support as it is now present for Devanagari short i matra.

Eric Muller favors 2. On the southasia list, 2008-11-13:

> The model that has been followed so far in Devanagari is to encode each vowel sign separately, even if they can be analyzed graphically as composites. For example, we have U+094B DEVANAGARI VOWEL SIGN O, and we don't use the sequence <U+093E DEVANAGARI VOWEL SIGN AA, U+0947 DEVANAGARI VOWEL SIGN E> to represent a vowel sign o.

> The current encoding proposal would break that model, by introducing a single new coded character, which could be used in combination with other coded characters for vowel signs. The question is whether we want to go down that path or whether we want to stick to the current model.

> Keeping the current model would suggest four coded characters

> devanagari vowel sign prishthamatra e
> devanagari vowel sign prishthamatra ai
> devanagari vowel sign prishthamatra o
> devanagari vowel sign prishthamatra au

> The last three coded characters would be "two part" vowels, which have received two treatments in Unicode: with a canonical decomposition (scripts of India) or without (Khmer). If I understand correctly, the canonical decomposition was primarily a concession to implementations and was not primarily intended to depart from the model of atomic vowel signs.

As Peter Constable pointed out, all three approaches suffer from confusables: काक would is either kaaka in modern orthography (<915, 93E, 915>) or kake in Prishthamatra orthography (<915, 915, 94e> under 1 and 2, and <915, 915, 947> under 3).

3.2 Rigvedic

accepted	Everson/Scharf			Govt. India		name(s)
accepted						name(s)

		07-343	08-050		07-39x	08-042	
	⌐̥	1cdc	1ce0	⌐̣		1ce7	vedic tone rigvedic kashmiri independent svarita vaidika uurdhva vakra rekhaa

Peter Scharf, on the southasia list, 2007-11-07:

Although Scharf and Everson provide ample evidence in European and Indian printed editions to justify the encoding of 1cdc (see n3366 figure 3, p. 14), manuscript evidence of the historical indigenous Indian use of the character is sought to allay Joshi's reservations that it may be a modern editorial invention.

Sobi, on the southasia list, 2008-01-03

In the Indian editions of Rigveda Khilani 1.11.4 and 1.12.7 (Poona edition), the sign is clearly depicting numeral '8' in Devanagari (southern-marathi numeral style) as an accent mark. This has been further commented and confirmed by Dr. B.B.Choubey in his 'Vedic Svarita Mimansa' (Hoshiyarpur Edition,1972 pg:107), unlike what is indicated in the documentary support under 3c and 3d in document L2/07-343.

There could be a possibility that 1CDC is based on North Indian-Hindi numeral '8', where the upper stroke of numeral '8' is horizontal as compared to the upper stroke of southern variety of numeral '8' which shows the same stroke angularly.

Therefore we propose the following:

Instead of giving separate code as suggested in L2/07-343 , the code point A8E8 can be modified to look like more or less horizontal '8', which could double up for 1CDC VEDIC TONE RIGVEDIC KASHMIRI INDEPENDENT SVARITA

Further, it must be noted that neither of the two proposals provide documentary support for this code A8E8. So now we are left with two alternatives:
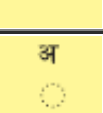
1.To delete A8E8 (L2/07-343), although it may cause an inelegant gap; and to have a separate code for 1CDC

2.Or to keep A8E8 with slight change in the form which can be effectively used as VEDIC TONE RIGVEDIC KASHMIRI INDEPENDENT SVARITA and also as combining Devanagari Digit Eight in Samaveda if need arises(since it is not yet found).

We would prefer the second alternative to encode 1CDC under Samaveda and this will be effected in our Revised Combined Proposal

Peter Scharf, on the southasia list, 2008-01-15

The 8 was a typographic substitute for what is shown in the mss. I have now provided the mss evidence in [L2/08-035].

3.3 Samavedic, letters and digits

accepted		Everson/Scharf		Govt. India			name(s)
		07-343	08-050		07-39x	08-042	
	०	a8e0	a8e0	O	8b4	a8e0	combining devanagari digit zero vaidika saamasvara anka shuunya
	१	a8e1	a8e1	१	8b5	a8e1	combining devanagari digit one vaidika saamasvara anka eka udatta
	२	a8e2	a8e2	२	8b6	a8e2	combining devanagari digit two vaidika saamasvara anka dvi svarita
	३	a8e3	a8e3	३	8b7	a8e3	combining devanagari digit three vaidika saamasvara anka tri anudatta
	४	a8e4	a8e4	४	8b8	a8e4	combining devanagari digit four vaidika saamasvara anka chatur
	५	a8e5	a8e5	५	8b9	a8e5	combining devanagari digit five vaidika saamasvara anka pancha
	६	a8e6	a8e6	६	8ba	a8e6	combining devanagari digit six vaidika saamasvara anka shatt
	७	a8e7	a8e7	७	8bb	a8e7	combining devanagari digit seven vaidika saamasvara anka sapta
	८	a8e8	a8e8	८		a8e8	combining devanagari digit eight vaidika saamasvara anka ashtta
	९	a8e9	a8e9	९		a8e9	combining devanagari digit nine vaidika saamasvara anka nava
	अ	a8ea	a8ea	अ	8bc	a8ea	combining devanagari letter a vaidika saamasvara abhinihita
	उ	a8eb	a8eb	उ		a8ee	combining devanagari letter u vaidika saamasvara u
		<a8e3, a8eb>	<a8e3, a8eb>	२उ	8c5	a8f0	vaidika saamasvara svarita dviu
							combining devanagari letter ka

क ◌	a8ec	a8ec	क ◌		a8ed	vaidika saamasvara ka
	<a8e3, a8ec>	<a8e3, a8ec>	इक ◌	8c6	a8f1	vaidika saamasvara svarita trika
न ◌	a8ed	a8ed	न ◌	8bd	a8eb	combining devanagari letter na vaidika saamasvara namana
प ◌	a8ee	a8ee	प ◌	8cb	a8f2	combining devanagari letter pa vaidika saamasvara prannatam
र ◌	a8ef	a8ef	र ◌	8bf	a8ec	combining devanagari letter ra vaidika saamasvara ra
	<a8e1, a8ef>	<a8e1, a8ef>	१र ◌	8c0	a8ef	vaidika saamasvara svarita ekara
	<a8e2, a8ef>	<a8e2, a8ef>	२र ◌	8c1	removed	vaidika saamasvara dvi ra
	<a8e3, a8ef>	<a8e3, a8ef>	३र ◌	8c2	removed	vaidika saamasvara tri ra
	<a8e4, a8ef>	<a8e4, a8ef>	४र ◌	8c3	removed	vaidika saamasvara chatur ra
	<a8e5, a8ef>	<a8e5, a8ef>	५र ◌	8c4	removed	vaidika saamasvara pancha ra
वि ◌	a8f0	a8f0	वि ◌	8ca	a8f3	combining devanagari letter vi vaidika saamasvara vinata
ऽ ◌	a8f1	a8f1	ऽ ◌	8c8	a8f4	combining devanagari sign avagraha vaidika saamasvara dirgibava

Even though the two proposals have the same approach and have considerable overlap, no characters in this group were accepted by the UTC at its October 2007 meeting, because there was no consensus on whether combining characters or ruby annotations are the best approach.

Note also that L2/07-343 proposes to use sequences where L2/07-39x proposes atomic characters.

Eric Muller and Peter Scharf, on the southasia list, 2007-11-08:

[Eric] I read the note http://sanskritlibrary.org/VedicUnicode/SLTN1.pdf which discusses ruby. As far as I can tell, the arguments are:

- unlike the general case of ruby in Japanese where the ruby base can be multiple characters/syllables, here the ruby base is always an akshara, so there is no scoping issue (and that scope is similar to that of other combining marks such as the vowel signs)

- the set of strings which appear as ruby text is a small, fixed repertoire

Here are a few counter arguments:

- looking the examples, I can't help but see the very mechanism of ruby at work, i.e. a mechanism of inter linear annotations rather than modifications of letters (which is what combining marks are generally used for)

- since the Jaiminiya tradition clearly requires a ruby implementation, an argument for combining marks based on availability of implementations is also one that suggests that the Jaiminiya tradition is a second class citizen

> [Peter]: No. The Jaiminiya notation has scoping issues that warrant Ruby just as Japanese does. These issues do not apply to the Kauthuma and Ranayaniya annotation methods.
>
>> [Eric]: My point is that if you go for combining marks because you expect implementations to be more readily available for combining marks than for ruby, and since Jaiminiya requires ruby, then you are willing to leave Jaiminiya behind, i.e. treat it as a second class system.

- the fact that there is no scoping issue and that combining characters would "work" is *not* an argument *against* ruby; it's only a necessary condition for combining marks

> [Peter]: Macrons, circumflexes, and accent marks in Roman could be handled by Ruby too, but they aren't.
>
>> [Eric]: But the accents in Latin do not write annotations, they are used to form new letters. One test is whether the text is still meaningful after you drop annotation/accents. If it is (as I suspect it is for the Vedic case), you have annotations; if it isn't (e.g., French), you have accents.

> [Peter] Nor are accent marks for Rgveda, Yajurveda, and Atharvaveda. For the same reasons that they aren't, such marks in Samaveda shouldn't be either.
>
>> [Eric] Which brings us to the distinction between interlinear annotations and non-interlinear annotations. I think it would be hard to argue that the neat arrangement we see in the examples is not interlinear.

- I suppose that one of the (unstated) arguments for combining marks is that implementations which can deal with, say, modern Hindi will just work with combining marks. Looking at an example like 4.1Qa in L2/07-343, I see that the placement of the 2nd and 3rd annotations, <digit 2> and <digit 1, vi>, is not strictly above their aksharas but spread to avoid collisions. This is a typical problem that ruby implementations have to deal with. I think that combining mark implementations will have

a hard time dealing with this correctly; stated another way, I am not convinced at all that an implementation for modern Hindi will work adequately. I fear that the attempt at getting something (Vedic support) for nothing is going to backfire.

- I suppose that the interest in the representation of Vedic text goes beyond the ability to display those texts, and includes the ability to manipulate them. It seems necessary for those manipulations to be able to reliably separate the annotated text from the annotations. In a combining mark representation, a candrabindu belonging to the annotated text and a digit one belonging the annotation have the same status (both are combining marks), which means that the characteristic of being an annotation is an integral part of the identity of the proposed characters. That seems a bit too much to me.

- the example to support a8e0/8b4 in L2/07-39x actually shows a latin 1 and a latin 2, if I am not mistaken. What would be used for those?

> [Peter] The Latin characters were employed in this example precisely because of the unavailability of Devanagari superscript characters such as those proposed. To meet such a need is the motivation for the current proposal.
>
>> [Eric] So either the example given in L2/07-39x as evidence for a superscript 0 is
>>
>> - not something anybody wants to be able to do with Unicode, and that proposal does not have evidence for a superscript 0 (no big deal at the end of the day, we have evidence from L2/07-343)
>>
>> - something that is wanted, but is not achieved by a superscript *devanagari* 0 (and the example says we would need combining *latin* 0, 1 and 2)
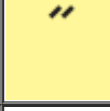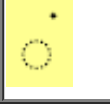
> [Eric] My main observation so far is that the arguments for one way or the other are not that clear cut, at least not at much as L2/07-343 pretends they are. With the arguments advanced so far, my opinion is in favor of ruby, but I would like to make sure that the best arguments are brought forward.

Following a question on the list, Eric Muller speculates that a ruby implementation is very likely to provide quite a bit of control over the styling of the ruby (e.g. font choice, point size, distance to the base character, top vs. above) and that a combining character implementation is unlikely to provide that control.

L2/08-050 reaffirms that ruby is entirely inappropriate and the combining characters are the best solution.
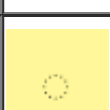
3.4 Samavedic, others

accepted	Everson/Scharf		Govt. India		name(s)
	07-343	08-050	07-39x	08-042	
					vedic tone karshana

		1cd1	1cd0		8c7	a8f5	vaidika saamasvara karshanna
		1cd2	1cd1		8a3	a8f6	vedic tone shara vaidika svarita uurdhva shara
		1cd3	1cd2		8c9	a8f7	vedic tone prenkha vaidika saamasvara prenkha
		1cd4	1cd3		8cc	a8f8	vedic sign nishshvasa vaidika saamagaan yogakaala
			1cd4		8be	removed	vedic sign kampa vaidika saamsavara kampa

See the comment for the previous section.

3.5 Yajurvedic, general

accepted	Everson/Scharf			Govt. India			name(s)
		07-343	08-050		07-39x	08-042	
y		1cd6	1cd6		8a8	1ceb	vedic tone yajurvedic independent svarita vaidika svarita adhah konna
y		1cd8	1cd7		8a9	1ce8	vedic tone yajurvedic kathaka independent svarita vaidika svarita adho vakra rekhaa
y		1cd7	1cd5		8a7	1cea	vedic tone yajurvedic aggravated independent svarita vaidika svarita adho nyubja
y		1cda	1cd8		8a6	1ce9	vedic tone candra below vaidika svarita adhah ardha vakra
		\<U+0951\>			8a0	removed	vaidika svarita uurdhva rekha
		\<U+0952, U+0952\> ?			8a1	removed	vaidika svarita adho dvi rekha
y		1cd4	1cda		8a4	1ce5	vedic tone double svarita vaidika svarita uurdhva dvi rekhaa
y		1cd5	1cdb		8a5	1ce6	vedic tone triple svarita vaidika svarita uurdhva tri rekhaa
							vedic tone dot below

y		1cdf	1cdd		8ac	1ced	vaidika svarita adho bindu
y		1ce0	1cdc		8ab	1cec	vedic tone kathaka anudatta vaidika svarita adho rekhaa
		1cd9	1cd9			1cee	vedic tone yajurvedic kathaka independent svarita schroeder vaidika svarita adho samyukt rekhaa

Note that U+0951 DEVANAGARI STRESS SIGN UDATTA and U+0952 DEVANAGARI STRESS SIGN ANUDATTA are already encoded, and that 0951 has the annotation "mostly used for Rigvedic svarita, with rare use for Yajurvedic udatta".

Sobi, on southasia list, 2007-12-31:

> 1. 08A1 Vaidika Svarita adho dvi rekha may not be addressed as two Anudatta as it has been identified as Svarita. Generally Shatapatabrahmana does not include svarita and has no sign for the same. However Weber in his edition of Shatapatabrahmana has a way to show independent Svarita by adding two horizontal lines to the prior Varna of the svarita Varna. e.g. Viryam

> To retain the authentic approach of Weber we have identified it as a Vaidika svarita Adho dvi Rekha and not as a two Anudatta, one below another.

> 2. In case of 1CD9 we are awaiting for the source material from which this sign has been identified. We are in process to examine parallel Indian Editions for the same.

Peter Scharf, on southasia list, 2008-01-15.

> George Cardona in his Bhasika System of Accentuation has demonstrated that the Satapathabrahmana marking indicates a two-tone system marking only anudatta. Although this two-tone sytem derives from the triple-tone system; it is not identical with it. The system designed by Weber needs to be reproduced but it is incorrect to identify double line below with a svarita mark. It marks an anudatta that precedes a svarita. Hence this should be produced by a sequence just as you agree the six dots should be produced by a sequence of two triple dot characters.

3.6 Yajurvedic, Satapathabrahmana

accepted		Everson/Scharf			Govt. India		name(s)
		07-343	08-050		07-39x	08-042	
y		1cdd	1cdf		8ae	removed	vedic tone three dots below vaidika svarita adhas tri bindu
		<1cdd, 1cdd>	<1cdf, 1cdf>		8af	removed	vaidika svarita adhas shatt bindu
							vedic tone two dots below

| y | | 1cde | 1cde | | 8ad | removed | vaidika svarita adho dvi bindu |

Sobi, on southasia list, 2007-12-31:

In case of 08AF Vaidika Svarita Adhas Shatt Bindu in L2/07-396, We agree to combine <08AE, 08AE> (with positioning underneath)> as recommended in the document L2/07-271. This will reflect in our revised combined proposal.

| accepted | Everson/Scharf | | | Govt. India | | name(s) |
	07-343	08-050		07-39x	08-042	
	<0967, 0951, 0952>			8b0	removed	vaidika svarita hrasva kampa
	<0969, 0951, 0952>			8b1	removed	vaidika svarita deergha kampa
	<0967, 0951>			8b2	removed	vaidika svarita hrasva kampa adhorekha
	<0969, 0951>			8b3	removed	vaidika svarita hrasva deergha kampa adhorekha
		<0951, 0952>			1cef	vaidika kampa
					1cf0	vaidika svarita urasi rekhaa

Sobi, on southasia list, 2007-12-31:

1. 08B0 Vaidika Svarita Hrasva Kampa and 08B1 Vaidika Svarita Deergha Kampa have been recommended as an integrated sign in L2/07-396.

The Hrasva Kampa is of one maatra duration in which first half maatra duration can be attributed to, as part of Udatta(Uddatansh)and the second half maatra duration belongs to part of Anudatta(Annudattansh). Whereas Deergha Kampa has two maatra time duration in which the first half maatra duration belongs to Udatta(Udattansh) and remaining 3 half maatra duration as part of Anudatta (Anudattansh).

1A. Thus the svarita kampa is a concept of integrated sound pattern of svarita/Udattaunsh and Anudattansh. Hence their dual identity has to be acknowledged by the integrated relevant accent mark as observed in Rigveda (Shaakal) kampa swara.

2. As special cases of Hrasva and Deergha Kampa, are observed as follows:

2A. In Atharveda Shaunaka the Hrasva kampa is shown with small numeral one and the horizontal

line Anudattansh below it. However, the vertical line as a svarita/udattansh is kerned backward and appears on the immediate previous independent svarita. Similar situation is also observed in Taitriya samhita.

In case of Deergha kampa a different variation is also observed where the small numeral 3 gets the vertical line (svarita/udattansh) on the top and the horizontal line (Anudattansh) is kerned backward and appears on the immediate previous independent svarita.

2B. In Rigveda Kashmir part, Kampa; Hrasva or Deergha is indicated by small numeral 3 with horizontal line (Annudattansh) and no vertical line (svarita/udattansh) appears on the top.

2C. In Maitrayini samhita the Kampa is shown by small numeral 3. However, the lower horizontal line (Anudattansh) is kerned forward on the next immediate independent svarita varna which could be either short or long.

2D. In the edition of Maitrayini samhita by Schroeder only small numeral 3 is indicated and no horizontal line (Anudattansh) is kerned forward on the next immediate independent svarita varna.

From the above now we suggest the following: (These action points would be reflected in our Revised Combined Proposal for Feb 2008 UTC meeting.)

A. In case of 1, only one code with vertical line above and horizontal line below could be considered in place of 08B0 and O8B1 and small numerals 1 and 3 can be attached to it as separate codes. This will reflect in our revised combined proposal.

B. In case of 2, considering all variations, it may be better to compose them with respective codes at respective places. Therefore the two codes 08B2 and 08B3 as suggested in L2/07-396 will be dropped out in our revised combined proposal.

Peter Scharf, on southasia list, 2008-01-15:

The variety you describe in points 2A-2D can all be accounted for by using the existing numeral, udatta, and anudatta marks. Font makers can design precomposed glyphs that compose the items with smaller numerals.

3.7 Atharvavedic

accepted	Everson/Scharf			Govt. India			name(s)
		07-343	08-050		07-39x	08-042	
y	⌇∫	1cdb	1ce1	⟨∫	8a2	1ce4	vedic tone atharvavedic independent svarita vaidika svarita dvi vakra khanda

3.8 Ardhavisarga

accepted	Everson/Scharf			Govt. India			name(s)
					07-		

		07-343	08-050		39x	08-042	
	⊖	1ce1	1ce2	⊖	8aa	1cde	vedic sign visarga svarita vaidika madhyerekhaa
		<0903, 1ce1>	<0903, 1ce2>	○÷	898	removed	vaidika visarga madhyerekha
	ꓸ	1ce2	1ce3	ꓸ		1cdf	vedic sign visarga udatta vaidika dakshinnatah uurdhvaga
		<0903, 1ce2>	<0903, 1ce3>	○ꓸ	89b	removed	vaidika visarga dakshinnatah uurdhvaga
	ϛ		1ce4	ϛ		1ce1	vedic sign reversed visarga udatta vaidika vaamatah uurdhvaga
			<0903, 1ce4>	○ϛ	899	removed	vaidika visarga vaamatah adhoga
	℮	1ce3	1ce5	℮		1ce0	vedic sign visarga anudatta vaidika vaamatah adhoga
		<0903, 1ce3>	<0903, 1ce5>	○℮	89a	removed	vaidika visarga vaamatah adhoga
	ꓹ		1ce6	ꓹ		1ce2	vedic sign reversed visarga anudatta vaidika dakshinatah adhoga
	ꙅ		1ce7				vedic sign visarga udatta with tail
			<0903, 1ce7>	○ꙅ	89c	removed	vaidika visarga dakshinnatah uurdhva vakra
	ꙅ		1ce8				vedic sign visarga anudatta with tail
			<0903, 1ce8>	○ꙅ	89d	removed	vaidika visarga vaamatak adho vakra

L2/07-343 proposes the encoding of pieces while L2/07-39x proposes the encoding of combinations of pieces.

Peter Scharf, on southasia list, 2007-11-07:

> In their meeting on 1 November in Paris, Joshi concurred that encoding the accent marks separately from the visarga which they accent would provide these advantages and agreed with the proposal to encode them thus in spite of reservations regarding the difficulties font designers would encounter in positioning the 1ce3 to the left of the visarga.

Conversely, Scharf accepted the possibility that the flourishes that appear on Joshi and Irani's 089c and 089d may be contours that depict tonal melodic flourishes. Further evidence of this from Vedic recitation or texts describing it would strengthen the case for a character encoding of the more elaborate udatta and anudatta visarga accents separate from that of the simpler udatta and anudatta accents. Scharf and Everson figure 7Bb, in n3366 p. 24 shows the simple and the flourished udatta accents in red marking the underlying visarga in black in a manuscript. The fact that both occur in the same passage probably indicates that they are intended to mark a significant difference, and the fact that they are written in a different ink indicates that they are independent characters from the visarga that they embellish.

Ganti Shantosh, on the southasia list, 2007-12-12:

1) 08AA : It may be discussed under 3.5, being Svarita sign and semantically it is not associated with Ardhavisarga.

2) 0898,0899,089B,089A are semantically well associated with the visuals of two vertical dots (as Visarga) and hence we have recommended the separate codes for all of the above , integrated with the other curvilinear elements in 0899,089B & 089A.This will help to create the plain-text of Vedic Sanskrit without much complexity.

3) Further, it may be agreed that the composed versions for 0898, 0899, 089B, 089A with two codes might help in adding an extra colour(deep red),but it is not observed in any printed source material as revealed in the support documents provided.

4) This approach will add more complexities in terms of positioning these curves(0899,089A,089B) with no economical gains and further, in the font area default Visarga(0903) may have to be substituted with another Visarga of broader width to accommodate 0898,0899,089A,089B characters with a post context reference.

5) The case of 089C,0890: The supporting documents reflect the need for these two cases. However, if 089C & 089D is treated as a variant of 089B & 089A respectively (and endorsed by scholars) then we could reconsider the above two codes.

This may raise another issue of provision codes for variant-one and variant-two(at least),as suggested by Miachel Everson in the document Encoding vedic Accents dated 2000-04-22

This is for further discussion.

Peter Scharf, on the southasia list, 2008-01-15:

I feel very strongly that the svarita markings on visarga be kept as separate characters to enable character level color coding of accents, as is evidenced in manuscripts. The fact that printed texts did not capture this feature of free-hand writing does not justify continuing the inability to capture Vedic writing systems properly in the digital medium.

If evidence does show that 089C & 089D had a different significance, for instance that the tonal pattern is more complex, I would favor encoding the svarita elements separately from the visarga here as well.

accepted	Everson/Scharf		Govt. India			name(s)
	07-343	**08-050**		**07-39x**	**08-042**	
	⪥	1ce4	1cf1				vedic sign ardhavisarga
			◌✕	89e	removed	vaidika jihvaamuliya vajra
			◌⪥	89f	1ce3	vaidika jihvaamuliya upadhamaaniya

The UTC earlier approved *U+0C71 TELUGU SIGN ARDHAVISARGA (proposed in L2/06-250, same shape as 1ce4, but combining), but it was pulled from the ballot by WG2 when it was realized that it was probably better treated as a general Vedic sign, rather than as a specifically Telugu sign.

Peter Scharf, on southasia list, 2007-11-07:

> In their meeting in Paris 1 Nov., Joshi argued that a sharp-cornered wedge shaped half-visarga is distinct in shape from the rounded shape so that two characters 089e and 089f are justified. Since the Kannada signs are already encoded, Joshi proposes a new character with common properties rather than changing the Kannada sign's properties and the insertion of remarks concerning equivalence of 089e with Kannada 0cf1 and of 089f with 0cf2. Remarks of equivalence should be added under the Kannada characters as well. Scharf concurs that 1ce4 can be merged with 089f and that 089e is the proper shape for the jihvaamuliya, i.e. not an x but wedges approaching each other. He maintains no objection to Joshi's proposal for encoding the vajra and gajakumbha but notes that the latter shape is commonly used both for jihvamuliya and upadhmaniya, which perhaps should be indicated in a remark.

Ganti Shantosh, on the southasia list, 2007-12-12:
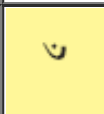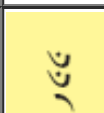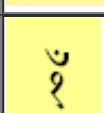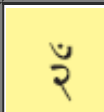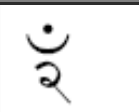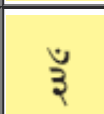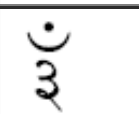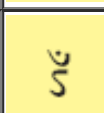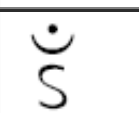
> The case of 089E & 089F: The two separate signs are semantically and visually two different signs which are proposed for one in the shape of Vajra(089E,angular weapon) and the other one as a Gajakumba(089F,variety of pot).Since, these two signs are associated grammatically with two different situations to grant them two separate codes is in order.

Peter Scharf, on the southasia list, 2008-01-15:

> I don't see evidence of the semantic distinction between vajra and gajakumbha on pp. 52-53 of L2/07-395,396,397.

3.9 Nasals

accepted	Everson/Scharf		Govt. India			name(s)
	07-343	**08-050**		**07-39x**	**08-042**	

		900	a8f2			900	devanagari sign inverted candrabindu vaidika adhomukha candrabindu
		a8f2	a8f3			1cd0	devanagari sign spacing candrabindu vaidika candrabindu
y		a8f3	a8f4		889	1cd1	devanagari sign candrabindu virama vaidika anusvaara candrabindu tiryak
y		a8f4	a8f5		88d	1cd5	devanagari sign double candrabindu virama vaidika anusvaara dvi candrabindu tiryak
			a8f6				devanagari sign candrabindu one
y		a8f5	a8f7		88a	1cd2	devanagari sign candrabindu two vaidika anusvaara candrabindu sadvi
y		a8f6	a8f8		88b	1cd3	devanagari sign candrabindu three vaidika anusvaara candrabindu satri
y		a8f7	a8f9		88c	1cd4	devanagari sign candrabindu avagraha vaidika anusvaara candrabindu saavagraha

Peter Scharf, on the southasia list, 2007-11-07:

> Joshi [...] wished to see evidence that 0900 inverted chandrabindu occurs in mss. and is not an editorial invention. Rosenfield has confirmed that it does occur in Kathaka mss. and is obtaining the evidence.

accepted		Everson/Scharf		Govt. India			name(s)
		07-343	**08-050**		**07-39x**	**08-042**	
		1ce5	1ce9				vedic sign anusvara antargomukha
		<1ce5, 0902>	<1ce9, 0902>		88e	1cd6	vaidika anusvaara antarmukha
			1ceb				vedic sign anusvara vamagomukha
			<1ceb, 0902>		88f	1cd7	vaidika anusvaara vaamamukha
			<1ceb,		890	1cda	vaidika anusvaara vaamamukha sacandra

		0901>	ँ			
౭	1ce6	1cea				vedic sign anusvaara bahirgomukha
	<1ce6, 0902>	<1cea, 0902>	ౣं	893	1cd9	vaidika anusvaara bahirmukha
	<1ce6, 0901>	<1cea, 0901>	ౣँ	894	1cdb	vaidika anusvaara bahirmukha sacandra
౭	1ce7	1cec				vedic sign vamargomukha with tail [EM: I suspect the "r" should be removed; see 1ceb]
	<1ce7, 0902>	<1cec, 0902>	ౙं	891	1cd8	vaidika anusvaara vaamamukha savakra

In all but one example, there is an anusvara or candrabindu component (there may be other signs as well). The only exception is example 8Ja of L2/07-343 which shows 1ce7 alone.

The approach proposed by L2/07-343 is to encode the unardorned signs, with the understanding that U+0902 DEVANAGARI SIGN ANUSVARA or U+0901 DEVANAGARI SIGN CANDRABINDU would be used as necessary. The basis is essentially to cut down on the number of characters to be encoded.

The approach proposed by L2/07-39x is to encode the combinations. The basis is that only the combinations are meaningful. The exception of example 8Ja is explained by:

> The Sign 1CE7 by itself is observed in some places but this could be because of composing errors as observed in "Maadyandina Asskhiyaama Satapathabraahmanam, Bharatiya Vidya Prakashan, Varanasi,1986,Page:7"where the same word repeated twice on the same page in lines 11 & 13. The word "havishi" is having 1CE7 with anudatta and 0891.

> To avoid such errors 0891 is needed or needed to be encoded.

Peter Scharf, on southasia list, 2007-11-07:

> In their meeting in Paris 1 Nov. Joshi argued that the difficulties of positioning multiple diacritics on 1ce5, 1ce6, and 1ce7, which never appear alone as such, justifies the encoding the greater number of characters with diacritics included. Scharf maintains no objection to Joshi's handling of these signs.

|	Everson/Scharf			Govt. India		name(s)
accepted	07-343	08-050		07-39x	08-042	

<0903, 1ce2, 1ce3>	<0903, 1ce4, 1ce6>	⦾	897	removed	vaidika anusvaara dvi bindu avagraha

Ganti Shantosh, on the southasia list, 2007-12-12:

Comment about 0897 : In L2/07-343 under figure 8Kb only the code 1CE8 & 0895 has been coded. However , the observation "The latter also shows VEDIC TONE VISARGA UDATTA and VEDIC TONE VISARGA ANUDATTA combined on a visarga in final position." has not been identified as a candidate for a code point. However the same sign has been found in "Sukla Yajurveda Samhitaa, Vasudev Laksmana Panasikara,Hindi vyakhyakar:Dr.Raamakrishna Saastri,Choukhamba VidyaBhavan,Varanasi,Page:102,Line:2"in a Padapaatha corresponding to Sahityapaatha on the same page(left column bottom) as an interpretation of Anusvaara sign 0893 and / or (1CE6,0902). Therefore the sign 0897 is presumed to be another type of Anusvaara appearing in Padapaatha. Hence, it is treated as vedic Anusvaara Dvi Bindu Avagraha and not Udatta Anudatta Visarga.

Peter Scharf, on the southasia list, 2008-01-15:

the shape occurs with moderate frequency to represent visarga with tonal mark. A single instance in which a presumption is made that it corresponds to nasalization is not sufficient evidence to override the frequent use it has as toned visarga. Scharf and Everson prefer not to encode the sign but to produce it with the combination of visarga + udatta + anudatta.

accepted	Everson/Scharf			Govt. India			name(s)
		07-343	08-050		07-39x	08-042	
	६		1cee	६	892	1cdc	vedic sign hexiform long anusvara vaidika anusvaara anugami
	६	1ce8	1cef		895	removed	vedic sign long anusvara
	६		1cf0	६	896	1cdd	vedic sign rthang long anusvara vaidika anusvaara tthasadrisha

Peter Scharf, on the southasia list, 2007-11-07:

On the http://sanskritlibrary.org/ site under Vedic Unicode proposal The document VedicMarks2007Mar10V.pdf Section C. Conclusions regarding Nasals, 1. Six dot is synonymous with rthang, presents evidence to show the synonymity of these characters. Six-dot (DAKSHINNAMUKHA) and Rthang (TTHASADRISHA) are typographic imitations of the VEDIC SIGN LONG ANUSVARA (CANCUMUKHA); all indicate the same phonetics in the same contexts.

Although Joshi concurs that the six dot and rthang (see L2/07-397, p. 44) are developments of 1ce8=0895, he argues that they deserve separate character status because their distinctive shapes are now commonly found, recognized, and referred to. In their meeting 1 November in Paris,

Scharf accepted Joshi's proposal to encode 0892 and 0896 as separate characters from 0895 cancumukha which is identical to Scharf and Everson's n3366: 1ce8.

Ganti Shantosh, on the southasia list, 2007-12-12:

The code point 0895 is observed in manuscript (Refer L2/07-397 document, Page no:44 Source Point No:4) and printed books as per document L2/07-397. In the same document L2/07-397 Page-43, the Source Point 1 and 2 shows the evidence for the code point 0895 itself.

Whereas, 0896 is the evolved form of 0895 for the print purpose in Northern editions. The horizontal line in this character has been Stressfully introduced to match with the top line of Devanagari script. As in the L2/07-397 document, Page no:44 Source Point No:4, one can observe the earlier handwritten version of this character.

It is also noted that such version was adopted in a typeface as seen in L2/07-397 document, Page no:44 Source Point No: 2 & 5. This version in our opinion is leading to the character code suggested under 0985 and 1CE8 in L2/07-343. This also establishes the fact that 0895 or 1CE8 is not the source of 0892.

In fact,"Maadyandina Asskhiyaama Satapathabraahmanam, Bharatiya Vidya Prakashan, Varanasi,1986,Page:7" clearly shows the separate and distinct usage of 0892,0895 and 0896.

Hence , the three separate codes as 0892,0895 & 0896 have been recommended in L2/07-397.

Peter Scharf, on the southasia list, 2008-01-15:

Scharf and Everson agree that 0896 developed as described from 0895 but disagree with the statement, "This also establishes the fact that 0895 or 1CE8 is not the source of 0892." On the contrary, 0892 is similarly a typographic imitation of 0895. We would have no objection to separately encoding these if some justification could be provided for distinguishing them as separate characters rather than as glyph variants. However, the three shapes occur without apparent difference. The synonymy of the boomerang and the six-dot is evident from the evidence in sections D2-3 of the document:
http://www.brown.edu/Departments/Classics/archive/Scharf/VajasaneyiAdditionalMarksV.pdf
(DRGG refers to Daulata Rama Gauda Gupta's edition of Vajasaneyisamhita) I am unable obtain the reference in order to verify the statement: "In fact,"Maadyandina Asskhiyaama Satapathabraahmanam, Bharatiya Vidya Prakashan, Varanasi,1986,Page:7" clearly shows the separate and distinct usage of 0892,0895 and 0896."

3.10 Additions for Devanagari

accepted	Everson/Scharf			Govt. India			name(s)
		07-343	08-050		07-39x	08-042	
		<0905, 0881>		अँ	880	removed	indo iranian avesta letter sarekah candra long a

y	◌ॕ	955	955	◌ॕ	881	removed	devanagari vowel sign candra long e indo iranian avesta vowel sign sarekha candra long a
		<0906, 094B>		ओ	882	removed	indo iranian avesta letter long aaoo
				◌ो	883	removed	indo iranian avesta vowel sign long aaoo
y	ॹ	979	979	ॹ	884	removed	devanagari letter zha indo iranian avesta letter zh
y	य	97a	97a	य	885	979	devanagari letter heavy ya vaidika letter jjya
				ब्ह	886	97a	vaidika letter llha
	ꣳ	973	973	ꣳ		1cf1	devanagari sign pushpika
	ˇ	974	974	◌ˇ		1cf5	devanagari sign divider vaidika trutikaa

Peter Scharf, on the southasia list, 2007-11-07:

> Joshi accepts 973 and 974.

Ganti Santosh, on the southasia list, 2007-12-10:

> In Proposal submitted by Michael Everson and Peter Schraf et al. L2/07-095,-095R,-230,343 only one Avestan dependent vowel sign 0955 Devanagari Vowel sign Candra Long E was included. We had put forward our views of inclusion of the Independent Vowel Letters and other Letters in our Observation 4 (first point) in the document L2/07-388.

> It is observed that so far 15 vowels and 34 consonants have been identified for Avesta pronunciation and transcription in Devanagari Script. (source: Avesta Part I by Kanga and Sontakke, Vaidika Sanshodhan Mandal, Pune) In addition to Short E and Short O there is a vowel pronunciation ARRA as in 'arrackal' which needs attention. As well as there are some peculiar consonant sounds of KH, for which reserved points were kept in 0887 and 0888 in L2/07-395,-396,-397.

> Obviously this needs more detailed probe. Therefore we tend to agree with the remarks made by Dr Michael Witzel (in his mail dated 19th October 2007 to Peter Scharf further forwarded to Prof R K Joshi) and with Peter Scharf's opinion (as expressed in his mail dated 8th November 2007 to Prof R K Joshi) that it may be advantageous for us to move all these into another area by themselves or into a separate proposal.

Therefore we may not consider Avestan characters into our Revised proposal due for the upcoming UTC meeting in Feb 2008. This may cause repercussion in the accepted list of characters accepted during the October 2007 UTC meeting #113.

Peter Scharf, on the southasia list, 2008-01-14

Scharf and Everson concur that 0881 (= n3366 0955) and 0884 (= n3366 0979) require encoding, but not with the encoding of characters 0880, 0882, 0883 for the following reasons.

0880 can be produced by the sequence 0905 + 0881; 0881 is identical to n3366 0955.

0882 can be produced by the sequence 0905 + 0883; it would also be able to be produce by the sequence 0906 + 094B. It is undesireable to have the same glyph producible by more than one sequence. Therefore it is preferable not to encode 0883 but to produce 0882 just by the sequence 0906 + 094B.

Apparently the CDAC proposal wishes to apply the principles of ISCII that underly the initial encoding of Indic scripts to Devanagari transcription of Avestan and to encode Avestan characters under analogous duplicate forms, i.e. both stand alone vowel forms and combining vowel forms. It is questionable whether this line of encoding should be pursued.

There is no doubt about the necessity of the Avestan characters for which the CDAC proposal and Scharf and Everson n3366 are in agreement. It is therefore undersireable to withdraw them, although Scharf and Everson have no objection to them being moved into a separate area for the transcription of Avestan and other scripts into Devanagari.

3.11 Miscellaneous

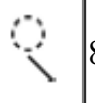| accepted	Everson/Scharf		Govt. India			name(s)
	07-343	08-050		07-39x	08-042	
			卐	8cd	1cf4	vaidika swastika

The UTC already approved four related characters, in particular 0FD5 TIBETAN SYMBOL GYUNG DRUNG NANG -KHOR, as documented in L2/07-148.

accepted	Everson/Scharf		Govt. India			name(s)
	07-343	08-050		07-39x	08-042	
			◌꣎	8ce	1cf2	vaidikka maatraakaal
			◌꣏	8cf	removed	vaidika apurrnnaanka ardha

Sobi, on southasia list, 2008-01-03:

08CE Vaidikka Apuurnnaanka Paada is graphically shorter as compared to danda and semantically is indicative of duration of certain time ie 1/4th of Maatraa Kaal

08CF Vaidika Apurrnnaanka Ardha is again graphically shorter as compared to double danda and semantically is indicative of duration of certain time ie 1/2 of Maatraa Kaal

| accepted | Everson/Scharf | | Govt. India | | name(s) |
	07-343	08-050	07-39x	08-042	
	◌		1ced	◌	8d0	1cf3	vedic sign tiryak vaidika tiryak

Peter Scharf, on the southasia list, 2007-11-07:

> Joshi argues that the diagonal stroke beneath gomukhas (see sec. 3.9 nasals esp. chars n3366 1ce5-1ce7 or Joshi 088e-0894, 0891) has a different history from the virama which is used only under consonant signs. The combining sign tiryak would allow for the composition of several glyph variants of nasal marks common in mss. and printed texts.

Sobi, on southasia list, 2008-01-03:

> 08D0 Vaidika Tiryak is not same as Viraama Virama (Halant) is used to suppress inherent vowel of a consonant where as Tiryak doesnt do so. Tiryak is used to indicate the qualitative aspect of svara of anusvara/anunasika. The word Tiryak is been used by Pandit. Satyavrath Shastri (Pg: xxi) in the introductory note in his book "Vaidika Vyakran - Translation of Vedic Grammar by Arthur A McDonald" (Motilal Banarasidas, 1971)

3.12 Jaiminiya

| accepted | Everson/Scharf | | Govt. India | | name(s) |
	07-343	08-050	07-39x	08-042	
				a8fe	vaidika prarambha chihna
				a8ff	vaidika antha chihna

4 Document history

- Wednesday January 30, 2008: folded L2/08-050 and L2/08-042/043 in.
- Saturday January 19, 2008: folded in the discussions on the southasia list up to now; made 3.1 more complete
- Monday November 19, 2007: pointing to local copies of the docs
- Monday November 5, 2007: typos fixed
- Monday October 29, 2007: first version