

Categorizing Some Uncategorized Indic Characters: A Report on UTC 164-A63

Roozbeh Pournader
2026-09-12

During UTC meeting #164, I was assigned an action item to “Re document L2/20-175 item F10, Investigate what the right Indic shaping properties should be for certain Devanagari and Vedic characters.” The action item was later passed down to SEW to track internally (#576) and was closed publicly. I have written this document to resolve that action item.

Proposal

I propose the following changes to the Indic_Syllabic_Category and Indic_Positional_Category properties:

- Change the Indic_Syllabic_Category property of U+1CE2..U+1CE8 VEDIC SIGN VISARGA SVARITA..VEDIC SIGN VISARGA ANUDATTA WITH TAIL and U+1CED VEDIC SIGN TIRYAK from Other to Cantillation_Mark.
- Change the Indic_Positional_Category property of U+1CF8 VEDIC TONE RING ABOVE and U+1CF9 VEDIC TONE DOUBLE RING ABOVE from NA to Top.

The original feedback

Date/Time: Tue Jul 7 17:43:56 CDT 2020

Name: Norbert Lindenberg

Report Type: Error Report

Opt Subject: Missing Indic shaping properties for Devanagari and Vedic characters

A number of Devanagari and Vedic characters are missing Indic syllabic or positional category definitions in the Unicode 13.0 data:

- 0950, 0971, A8F4..A8F7, A8FB, A8FD don't have an Indic syllabic category (as letters they don't need a positional category).
- 1CE2..1CE8, 1CED don't have syllabic categories (they do have positional categories).
- 1CF8..1CF9 don't have a positional category (they do have a syllabic category). For the first set, I can imagine that some of the characters don't participate in forming Devanagari syllables, and therefore the default value Other for the syllabic category is actually correct. If that's the case, however, I think it would be preferable to explicitly provide the value, both to make clear that it's intentional and to remind users of the data that this value can occur with Brahmic scripts (the specification of the Universal Shaping Engine currently does not handle this case).

U+0950 Devanagari Om and U+A8FD Devanagari Jain Om

We should retain InSC=Other. These represent the complete OM sign. None of the existing categories provides a better fit. Keeping InSC=Other also matches the Gujarati, Tamil, Sharada, Grantha, Newa, and Tirhuta OM signs.

U+0971 Devanagari Sign High Spacing Dot

This is already addressed in another recent document I submitted (SEW #931), where I recommend keeping InSC=Other.

U+A8F4..U+A8F7 “Marks of nasalization”

This is already addressed in another recent document I submitted (SEW #931), where I recommend changing their Indic_Syllabic_Category to Bindu.

U+A8FB Devanagari Headstroke

We should retain InSC=Other. Its documented function is to represent an illegible *whole orthographic syllable*, with repeated headstrokes indicating the extent of a *lacuna*. I found no documented use of HEADSTROKE as a base for dependent vowel signs or other combining marks. So I found no basis for changing its category.

U+1CE2..U+1CE8, “Diacritics for visarga”

We should change their Indic_Syllabic_Category to Cantillation_Mark. The Core Spec explicitly describes U+1CE2..U+1CE8 as combining marks indicating the tone of a visarga. Its example is <0903, 1CE2> for a svarita visarga. They fit Cantillation_Mark. Their function closely matches existing members of Cantillation_Mark. Their tone families have clear counterparts among the existing members:

Visarga marks	Tone family	Existing Cantillation_Mark comparison
U+1CE2	Svarita	U+1CD4 VEDIC SIGN YAJURVEDIC MIDLINE SVARITA
U+1CE3, U+1CE4, U+1CE7	Udatta	U+A8E1 COMBINING DEVANAGARI DIGIT ONE, used for udatta in Samaveda-Samhita
U+1CE5, U+1CE6, U+1CE8	Anudatta	U+0952 DEVANAGARI STRESS SIGN ANUDATTA; U+A8E3 COMBINING DEVANAGARI DIGIT THREE in Samaveda-Samhita

Attachment to visarga is not unprecedented within the category either. In its discussion of combining Devanagari digits and letters, [L2/07-343, p. 5](#), explains that final consonants, visarga, and certain nasals can become syllabified in Samaveda recitation and consequently carry their own tone marks. This means that the existing cantillation marks already extend beyond accents on ordinary syllables.

Three alternatives I don’t recommend are:

- InSC=Tone_Mark: functionally plausible, but would separate these from the existing Vedic accent marks without a compelling reason.
- InSC=Visarga: it would classify the accent component as though it represented the visarga itself.
- InSC=Syllable_Modifier: less specific than the established category for recitation marks.

These characters’ InPC=Overstruck value would not be new among InSC=Cantillation_Mark characters. U+1CD4 VEDIC SIGN YAJURVEDIC MIDLINE SVARITA already has InSC=Cantillation_Mark and InPC=Overstruck.

(There is also an implementation precedent: HarfBuzz’s USE generator already overrides this exact range to Cantillation_Mark.)

U+1CED Vedic Sign Tiryak

This character is hard to figure out. TIRYAK is well documented as the lower oblique stroke in certain Vedic nasal signs. Its independent phonetic effect is much less clear. I found no reliable rule saying that adding TIRYAK consistently changes a nasal’s pitch, duration, or articulation.

The Core Specification’s account is primarily graphical. It describes TIRYAK as a combining character used to render variants of nasal marks in manuscripts and printed texts. NamesList places it under nasalization marks and supplies the alias “vaidika tiryak” without further explanation. Its current properties are InSC=Other, InPC=Bottom, and ccc=220.

The actual examples we have in the UTC registry show the character as a component of a composite nasal sign:

Script	Documented combination
Devanagari	Gomukha nasal sign, upper dot, and lower TIRYAK. L2/08-043, p. 36
Bengali	BAHIRGOMUKHA with a dot and TIRYAK. L2/17-098, p. 12 . (Shriramana Sharma subsequently identified the Bengali example as the anusvara in <i>pauṃsnyam</i> . He described the complete form as another way of writing the same anusvara represented by the Bengali Vedic nasal sign. L2/18-035, p. 1 .)
Newa	ANTARGOMUKHA with TIRYAK and either anusvara or candrabindu; VAMAGOMUKHA with TIRYAK and candrabindu. L2/25-154, pp. 4–5
Sharada	BAHIRGOMUKHA with TIRYAK and Sharada candrabindu. L2/25-154, p. 6

So the attested hosts we know about are always *gomukha* shapes with a nasalization mark above.

The original proposal ([L2/08-042, p. 16](#)) demonstrates TIRYAK and a separate svarita together under a *gomukha*. (What it calls VAIDIKA ANUSVAARA is now represented as <ANUSVARA BAHIRGOMUKHA, ANUSVARA>, and what it calls VAIDIKA SVARITA is now

represented as VEDIC TONE CANDRA BELOW). This is the most complex example we have for the character, showing the *candra* below TIRYAK:

HA_n + AA_{vs} + VAIDIKA ANUSVAARA_n + VAIDIKA TIRYAK + VAIDIKA SVARITA_n

ह + ा + ं + ् + ृ → हांृ

The encoding-era discussions (p. 19) don't help much either. Scharf's November 2007 message explains the need to compose nasal glyph variants. Sobi's January 2008 message distinguishes TIRYAK from a virama suppressing a consonant's inherent vowel, then says it indicates a "qualitative aspect of svara of anusvara/anunasika." That provides no specific pronunciation rule:

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)

There is also some explanation I found from Michael Witzel from 1983 (https://www.academia.edu/18428638/Anun%C4%81sika_in_medieval_Veda_tradition), which calls the stroke attached to the nasal symbol a virāma/halanta and interprets it as evidence that the symbol originally was "not meant to be pronounced as a separate syllable." (He also discusses later reinterpretation of the nasal symbol as a separate syllable.) He does not use the name TIRYAK:

tradition. The application of a Virāma (*halanta*) attached to the Anu. symbol still shows that the letter originally was *not* meant to be pronounced as a separate syllable. – These facts can now be compared with grammatical tradition.

I tried a wider internet search using GPT-6 Astra, including Hindi/Sanskrit and Bengali terminology, but it yielded little additional explanation of the isolated stroke. Accounts of pronunciations such as *gum* generally concerned the whole nasal symbol.

Altogether, what we know is that it's a combining component used with particular Vedic nasal signs, and it can go between the nasal sign and a *candra* below.

This combined usage is my only guideline on how to assign an Indic_Syllabic_Category. So I looked at what Universal Shaping Engine category could be assigned to the character to make it work there. VMBlw/VOWEL_MOD_BELOW looked like the right USE category. It allows more VMBlw characters after it, which *candra below* is. Looking at the derivation of VM/VOWEL_MOD, this allows InSC=Cantillation_Mark, Tone_Mark, or Bindu, all resulting in the same USE category. (HarfBuzz overrides TIRYAK to Tone_Mark internally, also producing VMBlw.) Considering that we do not know for certain that TIRYAK marks tone or additional nasalisation, I think Cantillation_Mark (which is broadly defined as "recitation marks") is the best fit until we know better.

U+1CF8 and U+1CF9 "Signs for Jaiminiya Sama Veda"

These are clearly InPC=Top signs that were simply missing from the data file. Microsoft's supplementary USE positional data also assigns Top to both characters.

Signs for Jaiminiya Sama Veda

1CF8		VEDIC TONE RING ABOVE
1CF9		VEDIC TONE DOUBLE RING ABOVE

Final Note

Finally, it should be noted that the current USE classification table explicitly includes InSC=Other under IND/BASE_IND (<https://learn.microsoft.com/en-us/typography/script-development/use#character-classification>), so the part of the feedback that says "the specification of the Universal Shaping Engine currently does not handle this case" no longer holds. The specification was updated in August 2020.

IndicPositionalCategory.txt

1CF8..1CF9 ; Top

IndicSyllabicCategory.txt

1CE2..1CE8 ; Cantillation_Mark

1CED ; Cantillation_Mark

2 lines