

Some Changes to Indic_Syllabic_Category: A

Report on UTC 158-A106

Roozbeh Pournader

2026-09-09

During UTC meeting #158, Liang Hai and Roozbeh Pournader were assigned an action item to “Investigate David Corbett's feedback of Nov 25 and Nov 29, 2018 on the Unicode version 12.0 beta.” The action item was later passed down to SEW to track internally (#480) and was closed publicly. Roozbeh Pournader has written this document to resolve that action item.

Proposal

I propose the following changes to the Indic_Syllabic_Category property and IndicSyllabicCategory.txt:

- Change the Indic_Syllabic_Category property of U+0ECC LAO CANCELLATION MARK from Other to Consonant_Killer.
- Change the Indic_Syllabic_Category of U+1CFA VEDIC SIGN DOUBLE ANUSVARA ANTARGOMUKHA from Consonant_Placeholder to Bindu.
- Change the Indic_Syllabic_Category of the following characters from Other to Bindu:
 - U+1CE9..U+1CEC VEDIC SIGN ANUSVARA ANTARGOMUKHA..VEDIC SIGN ANUSVARA VAMAGOMUKHA WITH TAIL
 - U+1CEE..U+1CF0 VEDIC SIGN HEXIFORM LONG ANUSVARA..VEDIC SIGN RTHANG LONG ANUSVARA
 - U+A8F4..U+A8F7 DEVANAGARI SIGN DOUBLE CANDRABINDU VIRAMA..DEVANAGARI SIGN CANDRABINDU AVAGRAHA
 - U+11448 NEWA SIGN FINAL ANUSVARA
 - U+114C5 TIRHUTA GVANG

- Update the comments above the Bindu entries in Indic_Syllabic_Category to:
 # Bindu (anusvara, candrabindu, and related nasal signs)
 # Includes both combining marks and spacing nasal signs encoded as
 # letters. Some of the spacing signs can serve as bases for additional
 # combining marks, such as nasalization or cantillation marks.
- Add a NamesList annotation to U+0ECC LAO CANCELLATION MARK that it
 “indicates an unpronounced base consonant”, with cross-references to U+0E4C and
 U+17CD.

The original feedback

Date/Time: Sun Nov 25 15:53:00 CST 2018

Name: David Corbett

Report Type: Public Review Issue

Opt Subject: PRI #389: Indic_Syllabic_Category of U+11A3A ZANABAZAR SQUARE
 CLUSTER-INITIAL LETTER RA

**“The CLUSTER INITIAL RA is positioned at the normal head height, not above it”
 (L2/15-337 p. 23) so its Indic_Syllabic_Category should be Consonant_With_Stacker
 instead of Consonant_Prefixed.**

Date/Time: Thu Nov 29 15:30:24 CST 2018

Name: David Corbett

Report Type: Public Review Issue

Opt Subject: PRI #389: Missing values for Indic_Positional_Category and
 Indic_Syllabic_Category

U+111CA SHARADA SIGN NUKTA should have Indic_Positional_Category=Bottom.

**U+11300 GRANTHA SIGN COMBINING ANUSVARA ABOVE should have
 Indic_Positional_Category=Top.**

U+103C MYANMAR CONSONANT SIGN MEDIAL RA and U+1171E AHOM

CONSONANT SIGN MEDIAL RA should have a new Indic_Positional_Category, Top_And_Bottom_And_Left.

U+0971 DEVANAGARI SIGN HIGH SPACING DOT should have Indic_Syllabic_Category=Modifying_Letter.

U+0ECC LAO CANCELLATION MARK should have Indic_Syllabic_Category=Consonant_Killer.

U+19DA NEW TAI LUE THAM DIGIT ONE should have Indic_Syllabic_Category=Number.

The Vedic anusvaras U+1CE9 through U+1CEC, U+1CEE through U+1CF1, and U+A8F4 through U+A8F7 should have Indic_Syllabic_Category=Bindu.

U+11448 NEWA SIGN FINAL ANUSVARA and U+114C5 TIRHUTA GVANG should also have Indic_Syllabic_Category=Bindu.

Dispositions

Several of the requested changes have already been implemented, as summarized below:

Character(s)	Current value	Fixed
U+11A3A Zanabazar cluster-initial RA	InSC=Consonant_With_Stacker	17.0
U+111CA Sharada nukta	InPC=Bottom	12.0
U+11300 Grantha combining anusvara above	InPC=Top	12.0
U+103C Myanmar medial RA U+1171E Ahom medial RA	InPC=Top_And_Bottom_And_Left	13.0
U+19DA New Tai Lue Tham digit one	InSC=Number	12.0

The remaining issues are discussed in the following sections.

U+0971 Devanagari sign high spacing dot

The feedback asks for changing its Indic_Syllabic_Category to Modifying_Letter. There is currently only one such character, U+0B83 ு TAMIL SIGN VISARGA, normally called *aytham*. *Aytham* occurs in native Tamil words, but its function relevant to this category is as a spacing prefix that modifies the pronunciation of a following consonant, particularly to represent foreign sounds in loanwords. The category's current definition is:

Reanalyzed letters not participating in the abugida structure, but serving to modify the sound of an adjacent vowel or consonant.

The information I could find about the Devanagari sign was limited to the original proposal, [L2/06-137](#). It was encoded as a spacing tone/phonation sign, and it definitely isn't a letter subsequently reinterpreted as a modifier. There is some very general similarity of function. The proposal specifies that the sign starts a new "cluster", prevents reph formation across it, and remains before a following pre-base vowel sign. The available evidence does not establish that another existing InSC value would represent this behavior better than Other. No other existing category seems appropriate for it either (the closest one may be Tone_Mark). As such, I think it's best to leave it as InSC=Other.

U+0ECC Lao Cancellation Mark

Unfortunately, neither the NamesList nor the Core Spec has much information about the Lao character. I had to try other sources, including some in Lao language.

- A bilingual entry in a Lao online dictionary (<https://thesaurus.la/entry/ກຳຣັ້ນ>) says: "The cancellation mark (X) placed over a final consonant to indicate it is silent, primarily used for loanwords from Pali and Sanskrit."
- Laophaen's 2011 grammar discussion (<http://laophaen.free.fr/Dico/spip.php?article50>) calls the sign ໄມ້ທັນທະຄາດ, commonly called ໄມ້ກຳຣັ້ນ. Its question-and-answer explanation says that the consonant bearing the mark above it is not to be pronounced. It also describes the sign as retained for copying older documents and unnecessary in ordinary contemporary Lao spelling.
- The same website's 2012 discussion of names (<http://laophaen.free.fr/Dico/spip.php?article76>) provides an actual retained spelling: ພູເພັດຊຣັ້ນທອງ. The page says that, although use of the mark ceased long ago, a person's chosen spelling must be respected.
- The Lao Script Generation Panel's 2017 ICANN proposal (<https://www.icann.org/en/system/files/files/proposal-lao-lgr-31jan17-en.pdf>) includes U+0ECC as *mai ka lan*, requires it to follow a final consonant, and supports its inclusion with a 1935

Lao grammar, reproduced as Figure 7 on p. 22. The resulting LGR (<https://www.icann.org/sites/default/files/packages/lgr/rz-lgr-5-lao-script-26may22-en.html>) retains that context rule.

- Richard Ishida’s orthography notes (<https://r12a.github.io/scripts/laoo/lo.html>) say that it “was previously used to indicate silenced consonants, but is now described as obsolete.”

The feedback asks for changing its Indic_Syllabic_Category to Consonant_Killer. There are two existing characters in this category, U+0E4C THAI CHARACTER THANTHAKHAT and U+17CD KHMER SIGN TOANDAKHIAT. They are similar in both function and shape to the Lao mark under discussion. (The character has also been used in Pali to suppress vowels. This is a secondary use apparently shared by U+0E4C THAI CHARACTER THANTHAKHAT.) As such, I support changing the character’s InSC to Consonant_Killer. We may also want to add a NamesList annotation explaining that it indicates an unpronounced base consonant, with cross-references to U+0E4C and U+17CD.

The Bindu characters

The feedback asks for the fourteen remaining characters to become InSC=Bindu. These are all spacing characters, and tend to indicate nasalization. All fourteen presently have InSC=Other.

The closest comparisons are the existing spacing letters assigned InSC=Bindu:

- U+A8F2 SPACING CANDRABINDU and U+A8F3 CANDRABINDU VIRAMA: Devanagari spacing nasal signs aligned with the headline. They belong to the same series as the proposed U+A8F4..U+A8F7.
- U+09FC Bengali, U+0D04 Malayalam, and U+1145F Newa LETTER VEDIC ANUSVARA: Script-specific Vedic anusvaras, distinguished from the ordinary anusvara in the same script.
- U+1135E GRANTHA LETTER VEDIC ANUSVARA and U+1135F ... VEDIC DOUBLE ANUSVARA: Specialized Vedic nasals, including a doubled form. U+1135E represents different anusvara realizations according to context and can carry its own accent marks ([L2/09-372](#), Section 4.1). (Thus, specialized pronunciation, doubling, and serving as a combining-mark base already occur within Bindu.)
- U+0C80 KANNADA SIGN SPACING CANDRABINDU: Used in Badaga orthography and in Vedic texts. For the latter, it’s a half-anusvara pronounced approximately [ŋ] ([L2/15-158](#), p. 3). This is a nasal segment, not necessarily nasalization of the preceding vowel.

- U+A873 PHAGS-PA LETTER CANDRABINDU: Represents a final nasal in Sanskrit mantras, but is written and encoded before the syllable's other letters. (Consequently, even post-syllabic encoding order is not universal among spacing Bindu characters.)
- U+16D40 KIRAT RAI SIGN ANUSVARA; U+16D41 ... TONPI: The first represents [n] and [ŋ]; the second marks vowel nasalization. Both are modifier letters (gc=Lm). They provide an example of the category covering both nasal segments and nasalized vowels.

For the proposed additions, the evidence breaks down as follows:

- U+A8F4 DOUBLE CANDRABINDU VIRAMA: Anusvara before a spirant at the beginning of a consonant cluster. A specialized member of the same family as existing A8F2/A8F3. Strong case for Bindu.
- U+A8F5 CANDRABINDU TWO; U+A8F6 ... THREE: A nasalized vowel prolonged to two or three morae, respectively. These combine nasalization and vowel duration. They convey more than A8F2/A8F3, but their nasal function is explicit. Good case.
- U+A8F7 CANDRABINDU AVAGRAHA: Marks anusvara. Its documented function places it with the spacing nasal signs; the "avagraha" element in its name is insufficient reason to classify it as ordinary avagraha. Strong case.
- U+1CE9..U+1CEC, the four gomukha signs. Specialized nasal notation; they can combine with anusvara or candrabindu. Nasal function is well supported, and they can also serve as bases.
- U+1CEE HEXIFORM LONG ANUSVARA; U+1CEF LONG ANUSVARA; U+1CF0 RTHANG LONG ANUSVARA: Three synonymous signs indicating long anusvara after a short vowel. Specialized nasal length fits comfortably beside existing Vedic nasal letters and Grantha's double anusvara. Strong case.
- U+1CF1 ANUSVARA UBHAYATO MUKHA: The Core Spec specifies both nasal use and visarga with a tonal mark. Predominant nasal function is not established.
- U+114C5 TIRHUTA GVANG: Represents nasalization; its proposal explicitly groups it with antargomukha and bahirgomukha. Strong functional case.
- U+11448 NEWA SIGN FINAL ANUSVARA: Represents a nasal sound, but also has punctuation uses. A later proposal ([L2/17-440](#)) explicitly identifies nasal representation as its main use. Supportable as Bindu.

The Devanagari info above comes from [L2/07-343](#), Section 8, particularly p. 9; the generic Vedic distinctions from Core Spec [Section 12.1.8](#); and the Tirhuta description from [L2/11-175R](#), Section 4.11.

From the above set, the only unclear case is about U+1CF1. [The record](#) shows two different interpretations:

- Ganti Shantosh inferred an anusvara interpretation from a correspondence between passages in a particular edition.
- Peter Scharf responded that the shape occurred moderately frequently as a toned visarga, while the nasal interpretation rested on a single presumed instance. Scharf and Everson preferred representing the shape with visarga plus tone marks.

accepted	Everson/Scharf		Govt. India			name(s)
	07-343	08-050		07-39x	08-042	
	<0903, 1ce2, 1ce3>	<0903, 1ce4, 1ce6>		897	1cf7	vaidika anusvaara ubhayato mukha

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 Lakshmana 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 Udattha 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

<http://www.unicode.org/~emuller/southasia/vedic/>[1/29/2018 10:05:07 PM]

encode the sign but to produce it with the combination of visarga + udatta + anudatta.

I recommend waiting for more evidence or an actual proposal before we categorize U+1CF1 as either a Bindu or Visarga. I don't think we have enough information now.

Additionally, the research on this feedback brought up another example. U+1CFA VEDIC SIGN DOUBLE ANUSVARA ANTARGOMUKHA currently behaves a lot like some of the nasal characters mentioned above. It is currently classified as Consonant_Placeholder. Its encoding proposal ([L2/17-117R](#)) describes it as a nasal sign that can serve as a base for another nasal sign. Its documented function is consistent with the existing spacing characters assigned to Bindu. It should be reclassified as Bindu. It should be noted that the

change from Consonant_Placeholder to Bindu would not affect U+1CFA's ability as a base for following nasalization marks. Universal Shaping Engine (USE) derives its internal classes from both InSC and General_Category, and InSC=Bindu with gc=Lo results in a USE class of BASE (<https://learn.microsoft.com/en-us/typography/script-development/use#character-classification>), which can take nonspacing anusvaras etc.

Finally, we can also expand the comment above the Bindu class to explain all of this better. It currently says:

```
# Bindu/Anusvara (nasalization or -n)
```

We can expand it to:

```
# Bindu (anusvara, candrabindu, and related nasal signs)
# Includes both combining marks and spacing nasal signs encoded as
# letters. Some of the spacing signs can serve as bases for additional
# combining marks, such as nasalization or cantillation marks.
```

IndicSyllabicCategory.txt

0ECC; Consonant_Killer

1CE9..1CEC; Bindu

1CEE..1CF0; Bindu

1CFA; Bindu

A8F4..A8F7; Bindu

11448; Bindu

114C5; Bindu

7 lines