

Request to encode South Indian CANDRABINDU-S

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2010-Oct-11

This is a request to encode chandrabindu characters in the Telugu, Kannada and Malayalam blocks. These are currently (as of Unicode 6.0) not present in those blocks whereas North Indian scripts all have their own chandrabindu-s. While the Telugu block has the character 0C01 TELUGU SIGN CANDRABINDU, it will be shown that this character is actually unique to Telugu and is *not* the cognate of the chandrabindu of the other Indic scripts (like the so-called TAMIL SIGN VISARGA properly called āytam is unique to Tamil and is not the cognate of the visarga of the other Indic scripts). Thus the regular chandrabindu as has been encoded for North Indian scripts must also be encoded for these South Indian scripts.

§1. Background

The Indic chandrabindu is used to denote nasality, mostly of a vowel, but sometimes of semi-vowels as well, as seen in the following samples from pp 132-133 of ref 1, which is a comprehensive Sanskrit grammar written about 300 years back:

त्वं करोषि । सय्यन्ता, संयन्ता । सव्वत्सरः, संवत्सरः । यल्लोकम्, यं लोकम्
अत्रानुस्वारस्य पक्षेऽनुनासिका यवलाः । १२६ मो राजि समः कौ । (द-

संख्यं स्यात् । कियँ ह्यः, किं ह्यः । किवँ ह्लयति, किं ह्लयति । किल्लँ ह्लादयति, किं
ह्लादयति । १२६ नपरे नः । (द-३-२७) नपरे हकारे परे मस्य नः स्याद्वा । किन्

There are also some rare forms of writing Vedic Sanskrit in which the chandrabindu is placed on non-nasal class consonants to denote a nasal release, so:

वृक्कँगात्, चरक्कँतुः, अगँने, घँजन्ति
याच्चँभा, जँभानम्
आट्टँणारः
पत्तँली, ग्रथँजाति, ओद्दँमन्, पराधँमाताः
आपँजोति, गृभँगाति

The above examples are all of Sanskrit shown in the Devanagari script. Devanagari and other North Indian scripts also use the chandrabindu in the writing of the North Indian languages like Hindi. They already have individual chandrabindu characters encoded in Unicode. The South Indian scripts – Tamil, Telugu, Kannada and Malayalam – however do not use the chandrabindu for the writing of the respective South Indian languages since such nasality as shown above is not part of the linguistic content of these languages. This is why chandrabindu characters have not been encoded for these scripts in Unicode.

However it should be remembered that apart from the Tamil script, all the other major South Indian scripts – Telugu, Kannada and Malayalam – are fully capable of and in fact have been heavily used in representing Sanskrit in the respective regions. In the Tamil region the Grantha script has been used for Sanskrit. Thus all these four scripts – Telugu, Kannada, Malayalam and Grantha – will need the chandrabindu in the representation of Sanskrit for texts such as the Devanagari text shown above.

While the Grantha script is getting its own chandrabindu character encoded at 11301 (vide L2/09-372, L2/10-265R etc), the other three scripts do not yet have chandrabindu characters in Unicode. Usage of the chandrabindu in these scripts has however been attested. This proposal hence requests the encoding of these characters for the writing of Sanskrit/Vedic in these scripts.

§2. 0C01 TELUGU SIGN CANDRABINDU is not really a chandrabindu

An important matter to be noted here is that the character 0C01 TELUGU SIGN CANDRABINDU currently present in the Telugu block is not really a chandrabindu corresponding to that of other Indic scripts. It is an ‘arasunna’ or half-circle, and has been so annotated in the Telugu code chart. The semantics of this arasunna are quite different from those of the chandrabindu as described above.

While the pan-Indic chandrabindu is used as a modifier of vowel/consonant characters to indicate nasality, the Telugu arasunna or ‘half-anusvara’ is a variant of the sunna or ‘full anusvara’, and is used to indicate a historically elided anusvara. I provide references from Telugu linguistic texts (refs 2 and 3) regarding this on the next page. While one source (ref 4) I was able to access states that the arasunna is “used for Sanskrit words” it did not provide any explanations, examples or evidence for this. The chandrabindu is consistently different from this arasunna anyhow, both glyphically and in the GC (Mn as against Mc). Thus the arasunna is not a chandrabindu.

to serve as a cover symbol for all of them. Since [○] occurs only before stops and affricates, its phonemic status is automatically determined, i.e., /m/ before labial stops, /n/ before dental stops and affricates, /ɳ/ before retroflex stops, and /ŋ/ before velar stops. But historically, these pre-consonantal nasals were all lost regularly after long vowels, but optionally after short vowels. So, a form like /ēṅgu/ of inscriptional Te. or even Pre-Te. is always /ēgu/ in literary Te.; but an older form /kalaṅgu/ to be disturbed, is either /kalagu/ or /kalaṅgu/. Old Telugu grammarians devised a name for this alternating value of [○] (Romanization ṁ) and called it ardhānusvāra or **arasunna** (literally, half-nasal or half-circle).⁸ In all the environments of homorganic nasals they wrote the symbol of pūrṇānusvāra or full circle [◌], since its phonetic value (i.e., presence or absence of articulation) is determined by the length of the preceding vowel; therefore forms written [ēṁgu] and [kalaṁgu] were phonetically and phonemically ēgu, kalagu/kalaṅgu, respectively. It was only from prosody that one could know whether [○] was meant to be articulated as a full nasal or not in a given stanza after short vowels. In conformity with the meaning of **arasunna** (half-circle) the Te. printing system of classical texts introduced [◌] as opposed to [○]. [◌] means that it is a free alternant of a full nasal after short vowels and has a zero alternant after long vowels. Throughout this work I have

Ref 3 p 3 (from main text and footnote)

Telugu verbal bases can be brought under two structural patterns: (a) Those consisting of three short syllables, e.g. *palralcu* 'to spread', *malralu* 'to return'; this class also includes forms in which the middle syllable had originally a nasal (homorganic with the following suffix-consonant) as part of it, but which was lost from popular use in the pre-classical period;⁸ (b) Those with two syllables, of which the first is

⁸ This is called in Telugu grammars *ardhānusvāramu* or *ara-sunna* (i.e. half-nasal), which is only a historical remnant of a lost nasal orthographically preserved. This preservation has helped the poets to revive its phonetic value when they need it for metrical exigencies and ignore it when not: even in the 9th and 10th century inscriptions written in poetry, we have the complete nasals being written without their phonetic value; if they were to be read as nasals, the metre would be disturbed (see *Epigraphia Indica* XV: 150). G. J. Somayaji (1947) has presented some useful historical data on this problem in the chapter *ardhānusvaratattvam*.

§3. Attestation for the chandrabinidu

In contrast with the above usage of the arasunna, which was accidentally labeled the Telugu chandrabinidu for Unicode, the *real* chandrabinidu of Telugu is to be found in Sanskrit/Vedic texts as seen below:

Ref 5 p 96

ॐ॑ लो॒क॑ज्ज॒म॒य॒त्वा । तेषा॑ स॒म॒न्व॑य॒ति । य॒ज॒मा॒नो॑ वा॒ ष॒क॑
 त॒सा॒लः । सु॒व॒ग॒ण॑ लो॒क॑ अ॒हा॒व॒सि॑यः ॥ २४ ॥ य॒द्वे॒क॑वा॒ल॒मा॒
 वा॒व॒सि॑ये॒ ज॒हा॑सि । य॒ज॒मा॒न॒मे॒व॒ सु॒व॒ग॑ लो॒क॑ज्ज॒म॒य॒ति ।
 य॒द्व॒स॒ते॒न॒ ज॒हा॑सि॒यात् । सु॒व॒ग॑ लो॒का॑द॒य॒ज॒मा॒न॒म॒व॒वि॒धे॑त् ।
 सु॒प्रि॒चा॒ ज॒हा॑सि॒ति । सु॒व॒ग॑ स्य॒ लो॒क॑स्य॒ स॒म॒प्ये॑ । य॒त्रा॒भ्यु॒
 द्ये॒त । दे॒व॒लो॒क॑म॒भि॒ज॒ये॒त् । य॒द्द॒क्षि॑णा॒ पि॒तृ॒लो॒क॑म् । य॒त्प्रि॒
 त्य॑त् ॥ २५ ॥ र॒थो॒ ग॑ं॒ सि॒ य॒ज्ज॑ं॒ हा॒न्यः । य॒दु॒द॒क्षे॑ । म॒नु॒ष्य॑
 लो॒क॑म॒भि॒ज॒ये॒त् । प्र॒त्वि॒ष्यो॒॑ ज॒हा॑त॒व्यः । ष॒क॑वा॒ल॒म् प्र॒त्वि॒ष्य॑
 त्त्वा॒वा॒प॒धि॒वी॒ अ॒नु॒ प्र॒त्वि॒ष्य॑त् । द्या॒वा॒प॒धि॒वी॒ ऋ॒त॒वः । ऋ॒त॒वा॑
 य॒जः । य॒ज॒य॒ज॒मा॒नः । य॒ज॒मा॒न॑ प्र॒जाः । त॒न्मा॒त्प्रि॑ति॒
 य॒

Ref 5 p 127

य॒दो॒द॒नः॑ । प्र॒मा॒मे॒व॒न्त॑ं॒ त्रि॑य॒ज्ज॒म॒य॒ति । सु॒क्लो॑का॒ र
 सु॒म॒ज्ज॑त् र॒ स॒त्य॑रा॒जा॒ ३॒ नि॒त्या॒वा॒ । अ॒शि॒मे॒व॒ता॒मा॒ ॥ ३० ॥

It is evident that the Telugu chandrabinu used for Sanskrit/Vedic is certainly not the arasunna and that it is instead the same kind of chandrabinu as seen in other Indic scripts. Thus a separate 'real' chandrabinu character should be encoded for Telugu.

Likewise for Kannada one has:

Ref 6 Vol 1 p 622 – Maṇḍala 1 / Sūkta 1 / Mantra 2

ಅ॒ಗ್ನಿಃ ಪೂ॑ರ್ವೇ॒ಭಿ॒ರ್ಯು॑ಷ್ಠಿ॒ರಿಡ್ಶೋ॑ ನೂ॒ತನೈ॑ರು॒ತ |
ಸ ಧ್ವೇ॑ವಾ ಏ॒ಹ ವ॑ಕ್ಷತಿ || ೨ ||

Ref 6 Vol 2 p 111 – 1/4/6

ಉ॒ತ ನಃ॑ ಸು॒ಭ॒ಗಾ ಅ॒ರಿ॒ವೋ॒ರ್ಚೀ॑ಯು॒ರ್ದಸ್ಮ॑ ಕೃ॒ಷ್ವಯಃ॑ |
ಸ್ಯಾ॑ಮೇ॒ದಿ॒ಂದ್ರಸ್ಯ॑ ಶ॒ರ್ಮಣಿ॑ || ೬ ||

Ref 6 Vol 2 p 319 – 1/8/5

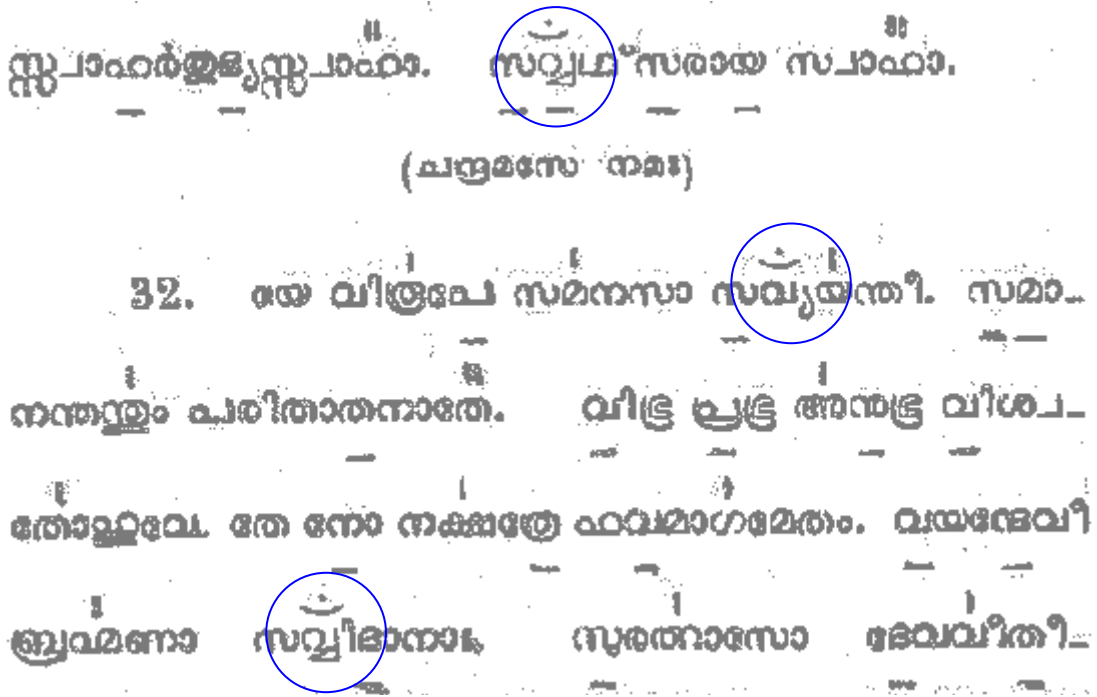
ಮು॑ಹಾ ಇ॒ಂದ್ರಃ ಪ॒ರಶ್ಚ॑ ನು ಮ॒ಹಿ॒ತ್ವಮ॑ಸ್ತು ವ॒ಜ್ರಿಣೇ॑ |
ದ್ಯೌ॑ರ್ನ ಪೃ॒ಥಿ॒ನಾ ಶ॒ವಃ॑ || ೫ ||

Ref 6 Vol 2 p 495 – 1/12/4

ತಾ॑ ಉ॒ಶ॒ತೋ॑ ವಿ ಬೋ॒ಧಯ॑ ಯದ॒ಗ್ನೇ॑ ಯಾ॒ಸಿ ದ್ರು॒ತ್ಯಂ |
ದೇ॒ವೈರಾ॑ ಸ॒ತ್ಸಿ ಬ॒ರ್ಹಿ॒ಷಿ || ೪ ||

Similarly for Malayalam one has:

Ref 7 p 66



§4. Unicode Character Properties

As there is already a character 0C01 TELUGU SIGN CANDRABINDU it is suggested that the character proposed here should be placed at 0C00 which is empty. The suggested name is ‘COMBINING CANDRABINDU ABOVE’ to distinguish it from the existing character. We note in passing that even in Devanagari there is an ‘INVERTED CANDRABINDU’ at 0900.

```

0C00;TELUGU SIGN COMBINING CANDRABINDU ABOVE;Mn;0;NSM;;;;N;;;;
0C81;KANNADA SIGN CANDRABINDU;Mn;0;NSM;;;;N;;;;
0D01;MALAYALAM SIGN CANDRABINDU;Mn;0;NSM;;;;N;;;;

```

An annotation should be added to 0C00 explaining that this is the ‘real’ chandrabinu equivalent to that of other Indic scripts. An annotation should also be added to the arasunna to the effect: “indicates a contextually elided nasal”.

§5. References

1. Siddhānta Kaumudī part 1, reprinted 1998 by Motilal Banarsidass Publishers, New Delhi
2. Telugu Verbal Bases: A Comparative and Descriptive Study, 1972, Bhadriraju Krishnamurti, <http://books.google.co.in/books?id=56PXtopfoscC&lpg=PA3&ots=0ZtKk3ydHf&dq=arasunna&pg=PA3>; this was apparently originally printed in 1961 in the University of California Publications in Linguistics, Volume 24, http://books.google.co.in/books?id=5glZAAAAMAAI&q=arasunna&dq=arasunna&hl=en&ei=-deVTIzAOYTKceCfGKQF&sa=X&oi=book_result&ct=result&resnum=4&ved=0CDcQ6AEwAw
3. Comparative Dravidian linguistics: Current perspectives, 2001, Bhadriraju Krishnamurti, <http://books.google.co.in/books?id=n9cnc0aUf-UC&lpg=PA27&dq=arasunna&pg=PA27>
4. Learn Telugu in 30 days through English, 2008, Krishna Gopal Vikal, <http://books.google.co.in/books?id=iPQw8lPjiREC&lpg=PA3&dq=arasunna&pg=PA3>
5. Taittirīya Brāhmaṇam, 1947, Ed: Brahmaśrī Naṭeśa Ghanapāṭhī, Pub: Gomaṭham Rāmānuja Jyautiṣika of the Mysore Maharaja Palace, Mysore.
6. Ṛg Veda Saṃhitā with Sāyaṇa Bhāṣya and Kannaḍa translation, Vols 1-32, 1947, Jayachamarajendra Veda Ratna Mala, Mysore.
7. Taittirīyopaniṣat, 2007 (4th edition), Ed: Paṇḍita Śrī La Anantarāmaśāstrī, Pub: R S Vadhyar and Sons, Kalpathi, Palghat.

§6. Proposed characters

0C00 ◌̣ TELUGU SIGN COMBINING CANDRABINDU ABOVE

0C81 ◌̣ KANNADA SIGN CANDRABINDU

0D01 ◌̣ MALAYALAM SIGN CANDRABINDU

§7. Official Proposal Summary Form

A. Administrative

1. Title

Request to encode South Indian CANDRABINDU-s

2. Requester's name

Shriramana Sharma

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution

4. Submission date

2010-Oct-11

5. Requester's reference (if applicable)

6. Choose one of the following: This is a complete proposal (or) More information will be provided later

This is a complete proposal.

B. Technical – General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters), Proposed name of script

No.

1b. The proposal is for addition of character(s) to an existing block, Name of the existing block

Telugu, Kannada and Malayalam.

2. Number of characters in proposal

3 (three)

3. Proposed category

Category B1 specialized.

4. Is a repertoire including character names provided?

Yes.

4a. If YES, are the names in accordance with the “character naming guidelines” in Annex L of P&P document?

Yes.

4b. Are the character shapes attached in a legible form suitable for review?

Yes.

5. Fonts related:

a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?

Not necessary. The required glyph may be copied from any other Indic block.

b. Identify the party granting a license for use of the font by the editors (include address, e-mail etc.)

Not necessary.

6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

No.

6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

Yes.

7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

Yes.

8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.

See detailed proposal.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

Ministry of IT, Govt of India. Vedic scholars of Andhra Pradesh, Kerala and Karnataka: Shri Dendukuri Narayana Haviryaji and Shri Dendukuri Shrirama Haviryaji of Tirupati for Telugu. Shri Hittilahalli Mahabaleshvara Bhat and Shri Amai Anantakrishna Ghanapathi of Bangalore for Kannada. Shri Ramesh Dravid of Palakkad, Kerala and Dr Sunandan Nambudiri of Chennai for Malayalam. Also Vinodh Rajan, a native speaker of Telugu with much experience in Indic transliteration.

2c. If YES, available relevant documents

None specifically. The matter was discussed in person or via telephone.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

These characters are used in the Telugu, Kannada and Malayalam scripts for the writing of Sanskrit/Vedic. Thus the user community comprises the Sanskrit/Vedic scholars of Andhra Pradesh, Karnataka and Kerala.

4a. The context of use for the proposed characters (type of use; common or rare)

Common in Vedic texts in Malayalam script. Somewhat rare in Vedic texts in Telugu/Kannada scripts.

4b. Reference

See proposal.

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

The Indian states of Andhra Pradesh, Karnataka and Kerala.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

6c. If YES, reference

They are analogous to chandrabindu characters of the other major Indic scripts in the BMP.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

No. Each character should be placed in the appropriate block of its script.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Yes, they are glyphically identical to the chandrabindu characters of other Indic scripts.

10b. If YES, is a rationale for its inclusion provided?

10c. If YES, reference

Their script property would differ. Indic scripts have their own chandrabindu characters.

11a. Does the proposal include use of combining characters and/or use of composite sequences?

Yes.

11b. If YES, is a rationale for such use provided?

11c. If YES, reference

The chandrabindu is by nature a non-spacing combining mark.

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

If YES, is the equivalent corresponding unified ideographic character(s) identified?

If YES, reference:

No.

-O-O-O-