

Proposal to encode the CANDRABINDU for Sinhala


Srinidhi A
srinidhi.pinkpetals24@gmail.com

Sridatta A
sridatta.jamadagni@gmail.com

February 5, 2018

1 Introduction

This is a proposal to encode the following character in the ‘Sinhala’ block in the BMP:

Glyph	Code	Character name
	U+0D81	SINHALA SIGN CANDRABINDU

2 Description

The Candrabindu is chiefly used to denote the nasalization of vowels and semivowels. It is attested in most Indic scripts including South Indian scripts. Candrabindu is distinctly encoded in Unicode in each script. It is observed that Sinhala Candrabindu is not encoded as nasalization is absent from Sinhala language.

The Sinhala script has full character repertoire to support Sanskrit and this script is commonly used for writing Sanskrit in Sri Lanka. The proposed character is used in Sanskrit texts such as Vyākaraṇa (Grammar) books.

The attestations are from Mugdhabodha Vyākaraṇa written by Vopadeva with a Sinhala translation Siddhāntaśekhāra by a distinguished scholar Weligama Sri Sumangala published in 1897, Colombo. ¹

Hence Candrabindu must also be encoded in Sinhala for representation of Sanskrit language.

¹ <https://archive.org/details/mugdabodhawyakar00vopauoft>

3 Character Data

Unicode character data: UnicodeData.txt

```
0D81; SINHALA SIGN CANDRABINDU;Mn;0;NSM;;;;;N;;;;;
```

Linebreak data: LineBreak.txt

```
0D81;CM # SINHALA SIGN CANDRABINDU
```

Indic syllabic categories: IndicSyllabicCategory.txt

```
# Indic_Syllabic_Category=Bindu
```

```
0D81; Bindu # Mn SINHALA SIGN CANDRABINDU
```

Indic positional categories: IndicPositionalCategory.txt

```
# Indic_Positional_Category=Top
```

```
0D81; Top # Mn SINHALA SIGN CANDRABINDU
```

Collation and other properties will be similar to that of Devanagari.

(මු-) 48. **ඡණ්ණං ඡණ්ණවතී ඡණ්ණගරීය:** (ඡජ-නාං ඡජ-නවතී-ඡටිනගරීය: 1...)

එතෙ නිපාතානතො*

(ස-) එතෙ, ඡණ්ණං ඡණ්ණවතී ඡණ්ණගරී යන මොව්හු ; නිපාතානතො, නිපාතනසෙන් සිඛකරණු ලැබෙත්. ඡජ-නාං—ඡකාරයහට මෙයින් ඡකාරව, (47) ඡණ්ණං, ඡටි-නව තී;—ටකාරයහට මෙයින් ඡකාරයි, (47) ඡණ්ණවතී; ඡටි-නගරී, මෙයින් ටකාරයහට ඡකාරයි, ඡණ්ණගරී.

(මු-) 49. **ලෙ ලජ්ජනා:** (ලෙ 7. ල: 1. තො: 6.)

ලෙ පරෙ තොලීකාර: සතත්. තලලය; විව්ලලිඛති.†

* අප්‍රධාන බැවින් යම් තැනෙක්හි නුමි ආගමය තොවත් නේද, එහි නිපාතනය නොමැත්තේය ; එහෙයින් අතිශයම කියාම වේ. ෂඩ්විකා නවතී: ඡණ්ණවතී: ඡණ්ණං නගරී ඡණ්ණගරී. නගරී යන සත්‍රීලිඛිතනිදේශය බැවින්; ෂඩ්වග රං යන මෙහි ඡකාරදෙශ නොවේ.

† අනුනාසිකය අභිවද්‍යාකාර වණියෙහි ප්‍රසිඛවෙයි, ඤම සංඥය සභාවයෙන් අනුනාසිකවෙයි, ඒ හෙතුවෙන් ඔහුගේ ස්ථානයෙහි හටගන්නාලබන යල් සංඥය සානුනාසික වෙයි —අභි වද්‍යාකාර වණී පුච්ඡම වන්නේය, කාරණ ගුණ: කායභිගුණ මාර්ගනො, යන නිකයෙන්ය යන අභිසී. මෙ සේ අනුඤාරයෙහි ජනවු යල්සංඥය සානුනාසික වෙයි.

විව්ලලිඛති යන මෙහි නිකාරයාගේ ස්ථානයෙහි ජනබැවින් ලකාරයහට අභි වද්‍යාකාර වණිය පුච්ඡවෙයි, අභිවද්‍යාකාර වණිය අනුඤාරය මෙන් පුච්ඡ හා සමබ්ධවෙයි. අර: යනු

Figure1. Use of Candrabindu in the word විව්ලලිඛති *vidvālikhati*.²

² In this document Candrabindu is marked in Latin transcriptions not in Sinhala due lack of font support.

(මු-) 52. වා ඛරසපෙ'රසං. (වා 1. තු 1. අරසපෙ 7. අරසං 1.)

දුනොන අදුනොන වා සභිතසා නො රෙඵචජියපෙ පරෙ රෙඵචජියමි වා සඤ්.†

* මෙහි පුඵසුත්‍රයෙන් සමිඛකරණ ලබන්නේද, නො: සි නැවත ගැන්ම සාධාරණවූ අනුඤ්ඤාය ගැණීම පිණිසයි. අන්පරිද්දෙකින් විනම් හුදකලා මකාර නකාරයන්ට ආදේශිත අනුඤ්ඤාය ගන්කල්හි, ජසඤ්ඤානො (827) යනාදිය සිදුනොවෙයි. ශාමිත: යනාදිය ශමධානුභව හා අන්කධානුභවද, කතයි.

† වෙන්වශයෙන් යොගය බැවින් අදුනොන යනු සමිඛකර නොවේ, එහෙයින් පදනතයෙහිද අපදනතයෙහිද මේ විධානයයි, එසේමැයි, පදනතයෙහි හෝ අපදනතයෙහි සිටි යාවූ අනුඤ්ඤායභව, යවල-යන මොවුන් පරකල්හි විකල්පයෙන්, යවල, යන මොව්හු වන්නාහුස, පදනතයෙහි සිටියා භව වනාහි වගීකරයා පරකල්හි තද්වගීනතයා විකල්පයෙන් වන්නේය යන අර්ථයි. පදනතයෙහිවූ අනුඤ්ඤායභව, යවල-යන මොවුන් පරකල්හි උදහරණ මෙසේයි, **ච්චය්චජ්ච** නිංයචජ්, **සච්චාණොමි**, සංචාණොමි, **ච්චලුබ්ධි: නිං ලුබ්ධි:** අපදනතයෙහි අනුඤ්ඤායභව, යවල-යන මොවුන් පරකල්හි මෙසේයි, **ය් ය්ම්ම්ම්ම්** යං යම්ම්ම්ම්, **ච්ච්ච්ච්ච්ච්ච්** චංචම්ම්ම්ම් **ල් ල්ල්ම්ම්ම්** ලංලම්ම්ම්ම්.

Figure 2. Use of Candrabindu in various instances as in the phrases ච්චය්චජ්ච *tvayyaccha*, සච්චාණොමි *saṅvrñomi*, ච්චලුබ්ධි: *tvalubdhah*, ය් ය්ම්ම්ම්ම් *ya ŷyamyate*, ච්ච්ච්ච්ච්ච්ච් *va ŷvamyate* and ල් ල්ල්ම්ම්ම් *la llamyate*.

(ස-) රෙඵච්ඡි යපෙ පරෙ, රකාරය රහිතවූ යජ්සංඥාපර කල්හි; දුනො, පදුනතයෙහි හෝ; වා, නොහොත්; අදු නො, පදුනත නුවූ තන්හි; ස්විතසූ, සිවියාවූ; නොෂ, අනු ඥායතං; වා, විකල්පයෙන්; රෙඵච්ඡියමි, රෙඵ රහිතවූ යමි සංඥා; සකත්, වන්වන්ය. [පාඨොති: 8. 4. 59.]

(මු-) යමො'නුනාසිකා: ය යයමානෙ, යං යමානෙ, හරි මහජ හරිංහජ. කමපනෙ ඉති පුච්ඡිණ නිත්තං. අරෙ කිං රංරමානෙ.*

(ස-) යමො'නුනාසිකා: යමි ප්‍රතිකාරය අනුනාසික වෙයි. යං-යමානෙ මෙයින් අනුඥායතං විකල්පයෙන් යකාර දෙයයි, ය යයමානෙ, පක්ෂයෙහි, යං-යමානෙ. හරිං-හජ මෙයින් අනුඥායතං විකල්පයෙන් මකාරයි, හරිමහජ, පක්ෂයෙහි, හරිංහජ. කන්-පනෙ (50) නකාරයහට අනුඥා රයි, කංපනෙ, (51) අනුඥායතං මකාරයි, කමපනෙ. ඉති පුච්ඡිණ නිත්තං, කමපනෙ යනු පුච්ඡිණයෙන් නිත්තයෙන් මකාරයි. අරෙ කිං, අරෙ යනු කුමක් සඳහාද? රංරමානෙ, යන මෙහි ර පරකල්හි ය ආදේශ නොවන බව ඇඟවීම පිණිසයි.

(මු-) 53. දුනො මො'සමන්තො හසෙ නු: (දුනො 7. ම: 6. අසමාප: 6. හසෙ 7. නු: 1.)

දුනො ස්විතසූ මසූ හසෙ පරෙ නු: සකත්, නතු සමාප: ශිවං සෙතාති, අසමාප: කිං, සමාපිති.†

Figure 3. Use of Candrabindu in the phrase ය යයමානෙ ya ÿyamyate.

ISO/IEC JTC 1/SC 2/WG 2

PROPOSAL SUMMARY FORM TO ACCOMPANY SUBMISSIONS

FOR ADDITIONS TO THE REPERTOIRE OF ISO/IEC 10646³.

Please fill all the sections A, B and C below.

Please read Principles and Procedures Document (P & P) from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/principles.html> for guidelines and details before filling this form.

Please ensure you are using the latest Form from <http://std.dkuug.dk/JTC1/SC2/WG2/docs/summaryform.html>.

See also <http://std.dkuug.dk/JTC1/SC2/WG2/docs/roadmaps.html> for latest Roadmaps.

A. Administrative

1. Title:	<i>Proposal to encode the CANDRABINDU for Sinhala</i>
2. Requester's name:	<i>Srinidhi A and Sridatta A</i>
3. Requester type (Member body/Liaison/Individual contribution):	<i>Individual contribution</i>
4. Submission date:	<i>February 5, 2018</i>
5. Requester's reference (if applicable):	
6. Choose one of the following:	
This is a complete proposal:	<i>Yes</i>
(or) More information will be provided later:	

B. Technical – General

1. Choose one of the following:	
a. This proposal is for a new script (set of characters):	
Proposed name of script:	
b. The proposal is for addition of character(s) to an existing block:	<i>Yes</i>
Name of the existing block:	<i>Sinhala</i>
2. Number of characters in proposal:	<i>1</i>
3. Proposed category (select one from below - see section 2.2 of P&P document):	

³ Form number: N4502-F (Original 1994-10-14; Revised 1995-01, 1995-04, 1996-04, 1996-08, 1999-03, 2001-05, 2001-09, 2003-11, 2005-01, 2005-09, 2005-10, 2007-03, 2008-05, 2009-11, 2011-03, 2012-01)

A-Contemporary	<input type="checkbox"/>	B.1-Specialized (small collection)	<input checked="" type="checkbox"/>	B.2-Specialized (large collection)	<input type="checkbox"/>
C-Major extinct	<input type="checkbox"/>	D-Attested extinct	<input type="checkbox"/>	E-Minor extinct	<input type="checkbox"/>
F-Archaic Hieroglyphic or Ideographic	<input type="checkbox"/>			G-Obscure or questionable usage symbols	<input type="checkbox"/>

4. Is a repertoire including character names provided?

Yes

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

Yes

b. 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 required. May be copied from any other Indic script.

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

N/A

6. References:

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

Yes

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

Yes

7. Special encoding issues:

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

See the text of proposal.

8. Additional Information:

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. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization

related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database (<http://www.unicode.org/reports/tr44/>) and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

C. Technical - Justification

1. Has this proposal for addition of character(s) been submitted before?	<i>No</i>
If YES explain	
2. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?	<i>No</i>
If YES, with whom?	
If YES, available relevant documents:	
3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?	
Reference:	<i>Used for Sanskrit by Sinhala speakers.</i>
4. The context of use for the proposed characters (type of use; common or rare)	<i>rare</i>
Reference:	
5. Are the proposed characters in current use by the user community?	
If YES, where? Reference:	<i>Details not available.</i>
6. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?	<i>Yes</i>
If YES, is a rationale provided?	<i>Yes</i>
If YES, reference:	<i>Sinhala is in BMP.</i>
7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?	<i>N/A</i>
8. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?	<i>No</i>
If YES, is a rationale for its inclusion provided?	

If YES, reference:	
9. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?	No
If YES, is a rationale for its inclusion provided?	
If YES, reference:	
10. Can any of the proposed character(s) be considered to be similar (in appearance or function) to, or could be confused with, an existing character?	Yes
If YES, is a rationale for its inclusion provided?	Yes
If YES, reference:	<i>Yes, they are identical to the candrabindu characters of other Indic scripts. Their script property would differ. Indic scripts have their own candrabindu characters.</i>
11. Does the proposal include use of combining characters and/or use of composite sequences?	Yes
If YES, is a rationale for such use provided?	Yes
If YES, reference:	<i>Candrabindu is a combining character.</i>
Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?	N/A
If YES, reference:	
12. Does the proposal contain characters with any special properties such as control function or similar semantics?	No
If YES, describe in detail (include attachment if necessary)	
13. Does the proposal contain any Ideographic compatibility characters?	No
If YES, are the equivalent corresponding unified ideographic characters identified?	
If YES, reference:	