

Universal Multiple-Octet Coded Character Set  
 International Organization for Standardization  
 Organisation Internationale de Normalisation  
 Международная организация по стандартизации

**Doc Type: Working Group Document****Title: Proposal to add three characters for Malayalam to the BMP of the UCS****Source: Michael Everson****Status: Individual Contribution****Action: For consideration by JTC1/SC2/WG2 and UTC****Date: 2007-08-17**

This document requests three additional characters to be added to the UCS and contains the proposal summary form. The characters are as follows:

0D29 MALAYALAM LETTER NNNA  
 0D3A MALAYALAM LETTER TTTA  
 0D78 MALAYALAM LETTER CHILLU Y

The first two characters represent alveolar consonants, transliterated *n* and *t*. They are used in scholarly orthography which distinguishes the alveolar sounds. The first has an analogue in modern Tamil, U+0B29 TAMIL LETTER NNNA. The second is another innovation which was subsequently lost because it merged with U+0D31 MALAYALAM LETTER RRA; RRA is pronounced as alveolar *r* or alveolar *t* depending upon context in modern Malayalam orthography. The third character is a chillu-form of YA, which contrasts with YA killed by a visible VIRAMA.

**Unicode Character Properties**

0D29;MALAYALAM LETTER NNNA;Lo;0;L;;;;N;;;;;  
 0D3A;MALAYALAM LETTER TTTA;Lo;0;L;;;;N;;;;;  
 0D7A;MALAYALAM LETTER CHILLU Y;Lo;0;L;;;;N;;;;;

**Bibliography**

- Grünendahl, Reinhold. 2001. *South Indian scripts in Sanskrit manuscripts and prints: Grantha Tamil, Malayalam, Telugu, Kannada, Nandinagari*. Wiesbaden: Harrassowitz. ISBN 3-447-04504-3
- Gundert, H. 1872 (1999). *Malayalam-English dictionary*. New Delhi & Madras: Asian Educational Services. ISBN 81-206-0066-5
- Pedersen, Thomas T. 2005-09-30. *Transliteration of Malayalam*. Retrieved 2007-08-11 from <http://transliteration.eki.ee/pdf/Malayalam.pdf>
- Rajaraja Varma, A. R. 1917. *Keralapanineeyam*. Reprinted 2003. Kottayam: DC Books.

Figures

<i>Sonorants (conjuncts p. 99)</i>		ha	○○
ya	യ & യ്	ɭa	ຍ & ຍ
ra	ര	ɭa	ຣ
la	ല	ra	○
va	വ	[na	6D] ←

Figure 1. Sample from Grünendahl 2001 showing NNNA.

	ISO 15919 2000 <sup>(1.0)</sup>	UN 1977 <sup>(2.0)</sup>	ALA-LC 1997 <sup>(3.0)</sup>
ത	ta	ta	ta
ഥ	tha	tha	tha
ദ	da	da	da
ധ	dha	dha	dha
ന	na <sup>(1.4)</sup>	na	na
പ	pa	pa	pa
ഫ	pha	pha	pha
ബ	ba	ba	ba
ഭ	bha	bha	bha
മ	ma	ma	ma
യ	ya <sup>(1.5)</sup>	ya	ya
ര	ra <sup>(1.6)</sup>	ra	ra
ല	la	la	la
വ	va	va	va
ശ	śa	sha	śa
ഷ	ṣa	ṣha	ṣa
സ	sa	sa	sa
ഹ	ha	ha	ha
ള	ɭa	ɭa	ɭa
ഴ	ɭa	ɭa	ɭa
റ	ra <sup>(1.6)(1.7)</sup>	ra <sup>(2.1)</sup>	ra <sup>(3.2)</sup>
ൺ	E ṅa	—	—
ൿ	E ṭa	—	—

Figure 2. Sample from Pedersen 2005 showing NNNA and TTTA.

37. ഴ, ഴ എന്ന ഴ വർഗ്ഗമാകട്ടെ, കവർഗ്ഗചവർഗ്ഗാദികൾപോലെ സ്പർശങ്ങളിലുള്ള വർഗ്ഗംതന്നെ. ഴഖരം; ഴ അനുനാസികം. അതിഖര മൃദുഘോഷങ്ങൾ തമിഴിൽ ഇല്ലല്ലോ. അതിനാൽ ഖരവും അനുനാസികവും മാത്രമേ ഉള്ളൂ. ഈ വർഗ്ഗത്തിന് സ്മാനം ദന്തമൂലത്തിനും മുർദ്ധാവിനും മദ്ധ്യേ ഉള്ള (ഊനിന് അടുത്ത) പിൻഭാഗം എന്ന പ്രദേശം; അതിനാൽ ഈ വർഗ്ഗം വർണ്യം. അക്ഷരമാലയുടെ രചന ഉള്ളിൽനിന്നും വെളിയിലേക്കുള്ള സ്മാനക്രമപ്രമാണിച്ചു ചെയ്തിട്ടുള്ളതാകയാൽ വർണ്യമായ ഴ വർഗ്ഗത്തെ മുർദ്ധന്യമായ ഴവർഗ്ഗം കഴിഞ്ഞ് ദന്ത്യമായ തവർഗ്ഗത്തിനു മുൻപായി പഠിക്കേണ്ടതായിരുന്നു; തമിഴുവൈയാകരണന്മാരാകട്ടെ, സംസ്കൃതാക്ഷരങ്ങളെ എല്ലാം മുറയ്ക്ക് എടുത്തതിന്റെശേഷം ഴ, ഴ എന്ന ഈ വർഗ്ഗത്തെ ഉള്ളതിലും ഒടുവിൽ തള്ളിക്കളഞ്ഞുവെന്നേ ഉള്ളൂ.

Figure 3. Sample from Rajaraja Varma 1917 showing NNNA and TTTA.

b. Final.

ക k റ r ണ ñ മ് = m യ് y റ് = r ത് = l ഡ് = l ഴ് = l'

c. Followed by a Consonant.

ക	ക	ക	ക	ക	ക	ക
ക	ക	ക	ക	ക	ക	ക
ക	ക	ക	ക	ക	ക	ക
ക	ക	ക	ക	ക	ക	ക
ക	ക	ക	ക	ക	ക	ക

[ക yka ക്ക kka ക്ക kka (ക yk ക്ക kya ക്ക kya ക്ക kya).]

Figure 4. Sample from Gundert 1872 showing CHILLU Y.

നായ് nāy T.M.Tu.C.Dog; pl. നായ്കൾ (kinds ചെന്നു. a wolf, കടൽനായ്. a seal, നീർനായ്. an otter, കഴുനായ്. etc.). നായ്കൻറെ മകൻ വായ്. (abuse). നായ്കൻറെ മകൻ കിഴക്കുവായ്കൻ. കയറിയ വായ്കൻ ഇറക്കിയവായ്കൻ (huntg.). നായ്കളെ കയറിട്ട മക്കൾ VeY. dog-boys. കൈയിൽ ഇരിക്കും പണയത്തക്കൊട്ടുമുള്ള മുഖത്തക്കടിക്കുന്ന നായ്കൻ No. (euph. of drunkards). നായ്കൻറെ നായ്കൻ വൃദ്ധം കാട്ടാലു prov. നായ്കൻ 1. a hunter. 2. the lowest caste of jungle-dwellers & beggars, (also called നായ്കൻ as dog-eaters), ordered to retire 74 steps from high-castes; hence രണ്ടു നായ്കൻ കിഴക്കു മൂലം മരം MR. 3. N. pr. m. (— ക്കി fem.)

Figure 5. Sample from Gundert 1872 showing CHILLU Y in nāy ‘wolf’.

ചായ് cāy 1. Beauty = ചായൽ. 2. a portion of killed game (of the 18 മരം 2 are called അടയ്കയും ചായ്കയും, huntg.) 3. = ചായ്, ചായ്.

ചായ് cāya & ചായ്ക T. M. (C. ചായ് to show, Te. to stretch, C. ചായ് to lie down & ചായ് = ചായ്. 1. To incline, lean to one side.

Figure 6. Sample from Gundert 1872 showing YA with visible VIRAMA in cāy ‘beauty’.

The Chill Effect

http://paivakil.port5.com/writings/chill\_effect.shtml

CPCN CO C24 KD Kw TV ET Blogs cP Fog Pnlp EntIE WG2 SC2 Udata

## CHILLING EFFECTS OF THE CHILLU

### ENCODING PROBLEMS IN MALAYALAM.

The vexed problem of chillu encoding has been a serious road block in adoption of the Unicode standard for Malayalam computing. People have been very quick to dismiss the Chillu as mere half a consonant.

I am no linguist, and will, for the time being, define a chillu as a special form of five specific consonants in Malayalam, rendered in a particular way. (Update:- see below). Did I say five? Actually, there are seven. The Malayalam language has seen ever so many "script reforms" at the behest of the printing industry, which are better called "simplifications" and each reform has resulted in reduction of the number of characters used in the language. The two chillus no longer used were victims of such reforms. I will not blame such reforms because limitations of printing technology, and volume of a print run for an average printed book in the Malayalam Language simply did not justify the existence and costs of a large number (900+) of glyphs in the Malayalam Language. Leaving history to itself, we will now look at the five chillus in frequent use and one chillu whose glyph I have with me right now in the table below.

The first row is the correct unicode encoding for the chillu in the 2nd row. The 2nd row shows the character, and the proper chillu form will show only if you have a unicode encoded font with Malayalam characters installed. The third row shows the accurate rendering of the chillu. If you get the same character in column 2 as the image in column 3, chances are that you are running a Unicode compliant operating system.

Table 1 - The Chillus In Malayalam

Unicode Sequence	Malayalam character as text	Malayalam text as Image
u+0D25 + u+0D4D + u+200D ന്	ൻ	
u+0D30 + u+0D4D + u+200D റ്	റി	
u+0D32 + u+0D4D + u+200D ല്	ൽ	
u+0D33 + u+0D4D + u+200D ള്	ൾ	
u+0D23 + u+0D4D + u+200D ണ്	ണി	
u+0D15 + u+0D4D + u+200D ക് (Not frequently used)	കി	
u+0D2F + u+0D4D + u+200D ഴ് (Not in use)		(The glyph is not available with me at the moment)

Figure 7. Sample from Mahesh Pai discussing CHILLU Y. He is not able to display the character.

Row 0D: MALAYALAM

	0D0	0D1	0D2	0D3	0D4	0D5	0D6	0D7
0		ഐ	ഠ	ര	ീ		ഋ	൧
1			ഡ	റ	ു		നു	൩
2	ം	ഒ	ഘ	ല	ു		ൺ	൯
3	ഃ	ഓ	ണ	ള	ു		ൺ	൪
4		ഔ	ത	ഴ	ു			ൽ
5	അ	ക	ഥ	വ				ൻ
6	ആ	ഖ	ഭ	ശ	െ		ഠ	
7	ഇ	ഗ	ധ	ഷ	േ	ൣ	൧	
8	ഈ	ഘ	ന	സ	ൈ		൨	യ്
9	ഉ	ങ	ണ	ഹ			൩	൯
A	ഊ	ച	പ	റ്റ	ൊ		൪	ൺ
B	ഋ	ഘ	ഫ		ോ		൫	ൻ
C	അ	ജ	ബ		ൣ		൩	ൽ
D		ഡ	ഭ	്	്		൫	ൽ
E	എ	ണ	മ	ാ			൮	ൾ
F	ഏ	ട	യ	ി			൯	ൽ

G = 00  
P = 00

Row 0D: MALAYALAM

hex	Name
00	(This position shall not be used)
01	(This position shall not be used)
02	MALAYALAM SIGN ANUSVARA
03	MALAYALAM SIGN VISARGA
04	(This position shall not be used)
05	MALAYALAM LETTER A
06	MALAYALAM LETTER AA
07	MALAYALAM LETTER I
08	MALAYALAM LETTER II
09	MALAYALAM LETTER U
0A	MALAYALAM LETTER UU
0B	MALAYALAM LETTER VOCALIC R
0C	MALAYALAM LETTER VOCALIC L
0D	(This position shall not be used)
0E	MALAYALAM LETTER E
0F	MALAYALAM LETTER EE
10	MALAYALAM LETTER AI
11	(This position shall not be used)
12	MALAYALAM LETTER O
13	MALAYALAM LETTER OO
14	MALAYALAM LETTER AU
15	MALAYALAM LETTER KA
16	MALAYALAM LETTER KHA
17	MALAYALAM LETTER GA
18	MALAYALAM LETTER GHA
19	MALAYALAM LETTER NGA
1A	MALAYALAM LETTER CA
1B	MALAYALAM LETTER CHA
1C	MALAYALAM LETTER JA
1D	MALAYALAM LETTER JHA
1E	MALAYALAM LETTER NYA
1F	MALAYALAM LETTER TTA
20	MALAYALAM LETTER TTHA
21	MALAYALAM LETTER DDA
22	MALAYALAM LETTER DDHA
23	MALAYALAM LETTER NNA
24	MALAYALAM LETTER TA
25	MALAYALAM LETTER THA
26	MALAYALAM LETTER DA
27	MALAYALAM LETTER DHA
28	MALAYALAM LETTER NA
29	MALAYALAM LETTER NNNA
2A	MALAYALAM LETTER PA
2B	MALAYALAM LETTER PHA
2C	MALAYALAM LETTER BA
2D	MALAYALAM LETTER BHA
2E	MALAYALAM LETTER MA
2F	MALAYALAM LETTER YA
30	MALAYALAM LETTER RA
31	MALAYALAM LETTER RRA
32	MALAYALAM LETTER LA
33	MALAYALAM LETTER LLA
34	MALAYALAM LETTER LLLA
35	MALAYALAM LETTER VA
36	MALAYALAM LETTER SHA
37	MALAYALAM LETTER SSA
38	MALAYALAM LETTER SA
39	MALAYALAM LETTER HA
3A	MALAYALAM LETTER TTTA
3B	(This position shall not be used)
3C	(This position shall not be used)
3D	MALAYALAM SIGN AVAGRAHA (praslesham)
3E	MALAYALAM VOWEL SIGN AA
3F	MALAYALAM VOWEL SIGN I
40	MALAYALAM VOWEL SIGN II
41	MALAYALAM VOWEL SIGN U
42	MALAYALAM VOWEL SIGN UU
43	MALAYALAM VOWEL SIGN VOCALIC R
44	MALAYALAM VOWEL SIGN VOCALIC RR
45	(This position shall not be used)
46	MALAYALAM VOWEL SIGN E
47	MALAYALAM VOWEL SIGN EE
48	MALAYALAM VOWEL SIGN AI
49	(This position shall not be used)
4A	MALAYALAM VOWEL SIGN O
4B	MALAYALAM VOWEL SIGN OO
4C	MALAYALAM VOWEL SIGN AU
4D	MALAYALAM SIGN VIRAMA
4E	(This position shall not be used)
4F	(This position shall not be used)
50	(This position shall not be used)
51	(This position shall not be used)
52	(This position shall not be used)
53	(This position shall not be used)
54	(This position shall not be used)
55	(This position shall not be used)
56	(This position shall not be used)
57	MALAYALAM AU LENGTH MARK
58	(This position shall not be used)

hex	Name
59	(This position shall not be used)
5A	(This position shall not be used)
5B	(This position shall not be used)
5C	(This position shall not be used)
5D	(This position shall not be used)
5E	(This position shall not be used)
5F	(This position shall not be used)
60	MALAYALAM LETTER VOCALIC RR
61	MALAYALAM LETTER VOCALIC LL
62	MALAYALAM VOWEL SIGN VOCALIC L
63	<i>MALAYALAM VOWEL SIGN VOCALIC LL</i>
64	(This position shall not be used)
65	(This position shall not be used)
66	MALAYALAM DIGIT ZERO
67	MALAYALAM DIGIT ONE
68	MALAYALAM DIGIT TWO
69	MALAYALAM DIGIT THREE
6A	MALAYALAM DIGIT FOUR
6B	MALAYALAM DIGIT FIVE
6C	MALAYALAM DIGIT SIX
6D	MALAYALAM DIGIT SEVEN
6E	MALAYALAM DIGIT EIGHT
6F	MALAYALAM DIGIT NINE
70	MALAYALAM NUMBER TEN
71	MALAYALAM NUMBER ONE HUNDRED
72	MALAYALAM NUMBER ONE THOUSAND
73	MALAYALAM FRACTION ONE QUARTER
74	MALAYALAM FRACTION ONE HALF
75	MALAYALAM FRACTION THREE QUARTERS
76	(This position shall not be used)
77	(This position shall not be used)
78	<u>MALAYALAM LETTER CHILLU Y</u>
79	MALAYALAM DATE MARK
7A	<i>MALAYALAM LETTER CHILLU NN</i>
7B	<i>MALAYALAM LETTER CHILLU N</i>
7C	<i>MALAYALAM LETTER CHILLU RR</i>
7D	<i>MALAYALAM LETTER CHILLU L</i>
7E	<i>MALAYALAM LETTER CHILLU LL</i>
7F	<i>MALAYALAM LETTER CHILLU K</i>



## A. Administrative

1. Title

**Proposal for to add three characters for Malayalam to the BMP of the UCS.**

2. Requester's name

**Michael Everson**

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

**Individual contribution.**

4. Submission date

**2007-08-17**

5. Requester's reference (if applicable)

6. Choose one of the following:

6a. This is a complete proposal

**Yes.**

6b. More information will be provided later

**No.**

## B. Technical – General

1. Choose one of the following:

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

**No.**

1b. Proposed name of script

1c. The proposal is for addition of character(s) to an existing block

**Yes.**

1d. Name of the existing block

**Malayalam.**

2. Number of characters in proposal

**3.**

3. Proposed category (A-Contemporary; B.1-Specialized (small collection); B.2-Specialized (large collection); C-Major extinct; D-Attested extinct; E-Minor extinct; F-Archaic Hieroglyphic or Ideographic; G-Obscure or questionable usage symbols)

**Category A.**

4a. Is a repertoire including character names provided?

**Yes.**

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

**Yes.**

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

**Yes.**

5a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

**Michael Everson.**

5b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

**Michael Everson, Fontographer.**

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

**Yes.**

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. 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/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

**See above.**

## C. Technical – Justification

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

**No, though C. J. Cibu proposed the CHILLU Y to the UTC.**

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?

**C. J. Cibu. Naga Ganesan**

2c. 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?

**Malayalam users.**

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

**Rare.**

4b. Reference

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

**Yes.**

5b. If YES, where?

**Dictionaries, grammars, books on linguistics.**

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?

**No.**

6c. If YES, reference

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

**No.**

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?

**No.**

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

10c. If YES, reference

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

**No.**

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

11c. If YES, reference

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

**No.**

11e. If YES, reference

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)?

**No.**

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?