ISO/IEC JTC1/SC2/WG2 N3295

L2/07-283

2007-08-17

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

Individual Contribution Status:

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 p 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 aleolar 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;;;;
OD3A; MALAYALAM LETTER TTTA; Lo; 0; L;;;;; N;;;;
OD7A; 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.

Sonorants	(conjuncts p. 99)	ha	00
ya	D & (0)	ļa	^{த ஆ} இ
ra	(0)	ļa	9
la	ല	ŗa	0
va	Оl	na	ബ്വ

Figure 1. Sample from Grünendahl 2001 showing NNNA.

		:	ISO 15919 2000(1.0)	UN 1977 ^(2.0)	ALA-LC 1997 ^(3,0)
	ා		ta	ta	ta
	מו		tha	tha	tha
	ß .		da	da	da
	$\mathbf{\omega}$		dha	dha	dha
	m		na ^(1.4)	na	na
	ol ·		pa	pa	pa
	ഫ		pha	pha	pha
	តារ		ba	ba	ba
	ß		bha	bha	bha
	Ø		ma	ma	ma
	യ		ya(1.5)	ya	ya
	0		га ^(1.6)	ra	га
	ല		la	la	la
	വ		va	va	va
	00		śa	sha	śa
	വിഷ		şa	şha	şa
	m		sa	sa	sa
	ഹ		ha	ha	ha
	ള		ļa	ļa	ļa
	φ		la	<u>l</u> a	ļа
	0		<u>r</u> a ^{(1.6)(1.7)}	ra(2.1)	<u>r</u> a(3.2)
	ຄ	E	ŋa		_
*	ц	Е	<u>t</u> a		_

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

37. 4, ന എന്ന വർഗ്ഗമാകട്ടെ, കവർഗ്ഗചവർഗ്ഗാദികൾപോലെ സ്പർശങ്ങളിലുള്ള വർഗ്ഗംതന്നെ. പ്രഖരം; ന അനുനാസികം. അതിഖര മൃദുഘോഷങ്ങൾ തമിഴിൽ ഇല്ലല്ലോ. അതിനാൽ ഖരവും അനുനാസികവും മാത്രമേ ഉള്ളു. ഈ വർഗ്ഗത്തിന് സ്ഥാനം ദന്തമൂലത്തിനും മൂർദ്ധാവിനും മദ്ധ്യേ ഉള്ള (ഊനിന് അടുത്ത) പിൻഭാഗം എന്ന പ്രദേശം; അതിനാൽ ഈ വർഗ്ഗം വർത്സ്യം. അക്ഷരമാലയുടെ ചേന ഉള്ളിൽനിന്നും വെളിയിലേക്കുള്ള സ്ഥാനക്രമംപ്രമാണിച്ചു ചെയ്തിട്ടുള്ളതാകയാൽ വർത്സ്യമായ വർഗ്ഗത്തെ മൂർദ്ധന്യമായ ടവർഗ്ഗം കഴിഞ്ഞ് ദന്ത്യമായ തവർഗ്ഗത്തിനു മുൻപായി പഠി ക്കേണ്ടതായിരുന്നു; തമിഴുവൈയാകരണന്മാരാകട്ടെ, സംസ്കൃതാക്ഷര ങ്ങളെ എല്ലാം മുറയ്ക്ക് എടുത്തതിൻറശേഷം 4, ന എന്ന ഈ വർഗ്ഗത്തെ ഉള്ളതിലും ഒടുവിൽ തള്ളിക്കളഞ്ഞുവെന്നേ ഉള്ളു.

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

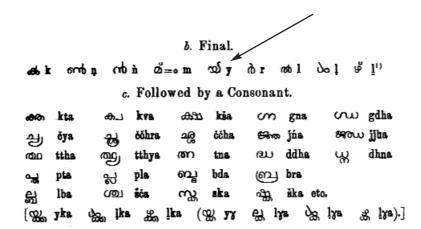


Figure 4. Sample from Gundert 1872 showing CHILLU Y.

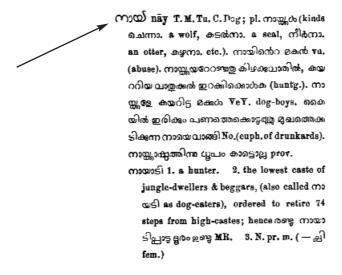


Figure 5. Sample from Gundert 1872 showing CHILLU Y in $n\bar{a}y$ 'wolf'.

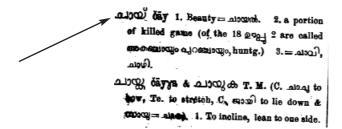
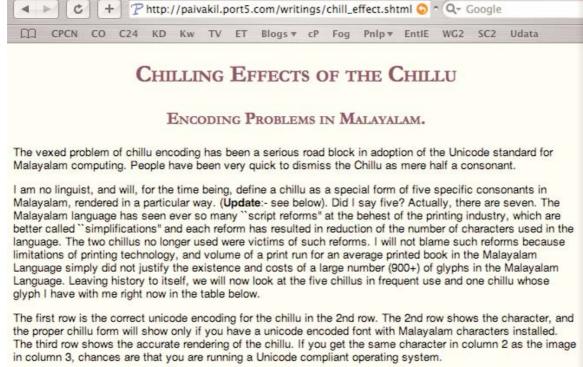


Figure 6. Sample from Gundert 1872 showing YA with visible VIRAMA in $c\bar{a}y$ 'beauty'.



The Chill Effect

Table 1 - The Chillus In Malavalam

Unicode Sequence	Malayalam character as text	Malayalam text as Image
u+0D25 + u+0D4D + u+200D	ന്	ൻ
u+0D30 + u+0D4D + u+200D	ര്	ô
u+0D32 + u+0D4D + u+200D	ല്	ൽ
u+0D33 + u+0D4D + u+200D	ള്	ൾ
u+0D23 + u+0D4D + u+200D	ണ്	ൺ
u+0D15 + u+0D4D + u+200D	க் (Not frequently used)	æ
u+0D2F + u+0D4D + u+200D	യ് (Not in use)	(The glpyh is not available with me at the moment)

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

00 00

Row 0D: MALAYALAM

	0D0	0D1	0D2	0D3	0D4	0D5	0D6	0D7	
0		ഐ	0	0	ീ		9	Ju	
1			w	0	ു		ആ	თ	
2	ം	в	w	ല	ൂ		୍ଦ	൲	
3		ഓ	ണ	<u>ള</u>	ൃ		ූ	σ	
4		ഔ	ത	g	ೃ			බ	
5	അ	ക	Ю	വ				൵	
6	ആ	ഖ	ß	S	െ		0		
7	ഇ	S	w	ഷ	േ	ာ	Ъ		
8	ഈ	ഘ	m	m	ൈ		Q	ගු	G = P =
9	ഉ	ങ	ഩ	ഹ			സ	൹	
Α	ഊ	ച	പ	Ц	ൊ		ф	ൺ	
В	8	ഛ	ഫ		ോ		<u> </u>	ൻ	
С	ത	ஐ	ബ		ൌ		m	ල	
D		ഝ	ß	ſ	ँ		ඉ	ൽ	
E	എ	ഞ	2	ာ			വ	ൾ	
F	ഏ	S	ω	ി			ൻ	<i>æ</i> 3	

Row 0D: MALAYALAM

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 A 07 MALAYALAM LETTER II 08 MALAYALAM LETTER II 09 MALAYALAM LETTER U 0A MALAYALAM LETTER U 0B MALAYALAM LETTER U 0B MALAYALAM LETTER U 0B MALAYALAM LETTER VOCALIC R 0C MALAYALAM LETTER VOCALIC R 0C MALAYALAM LETTER VOCALIC L 0D (This position shall not be used) 0E MALAYALAM LETTER E 0F MALAYALAM LETTER E 10 MALAYALAM LETTER AI 11 (This position shall not be used) 12 MALAYALAM LETTER O 13 MALAYALAM LETTER O 14 MALAYALAM LETTER O 15 MALAYALAM LETTER O 16 MALAYALAM LETTER O 17 MALAYALAM LETTER O 18 MALAYALAM LETTER O 19 MALAYALAM LETTER NA	
02 MALAYALAM SIGN ANUSVARA 03 MALAYALAM SIGN VISARGA (This position shall not be used) 05 MALAYALAM LETTER A 06 MALAYALAM LETTER I 08 MALAYALAM LETTER I 09 MALAYALAM LETTER II 09 MALAYALAM LETTER U 0A MALAYALAM LETTER U 0B MALAYALAM LETTER U 0C MALAYALAM LETTER VOCALIC R 0C MALAYALAM LETTER VOCALIC L (This position shall not be used) 0E MALAYALAM LETTER E 0F MALAYALAM LETTER E 10 MALAYALAM LETTER AI 11 (This position shall not be used) 12 MALAYALAM LETTER O 13 MALAYALAM LETTER O 14 MALAYALAM LETTER O 15 MALAYALAM LETTER O 16 MALAYALAM LETTER O 17 MALAYALAM LETTER O 18 MALAYALAM LETTER AU 19 MALAYALAM LETTER AU 10 MALAYALAM LETTER AU 11 MALAYALAM LETTER AU 11 MALAYALAM LETTER AU 11 MALAYALAM LETTER AU	
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 00 MALAYALAM LETTER U 01 MALAYALAM LETTER U 02 MALAYALAM LETTER VOCALIC R 03 MALAYALAM LETTER VOCALIC R 04 MALAYALAM LETTER VOCALIC L 05 MALAYALAM LETTER VOCALIC L 06 MALAYALAM LETTER E 07 MALAYALAM LETTER E 10 MALAYALAM LETTER E 11 (This position shall not be used) 12 MALAYALAM LETTER AI 13 MALAYALAM LETTER O 14 MALAYALAM LETTER O 15 MALAYALAM LETTER O 16 MALAYALAM LETTER O 17 MALAYALAM LETTER O 18 MALAYALAM LETTER AU 19 MALAYALAM LETTER AU 19 MALAYALAM LETTER AU 10 MALAYALAM LETTER AU 11 MALAYALAM LETTER AU	
06 MALAYALAM LETTER AA 07 MALAYALAM LETTER I 08 MALAYALAM LETTER II 09 MALAYALAM LETTER II 00 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 EI 11 (This position shall not be used) 12 MALAYALAM LETTER O 13 MALAYALAM LETTER O 14 MALAYALAM LETTER O 15 MALAYALAM LETTER OU 16 MALAYALAM LETTER OU 17 MALAYALAM LETTER OU 18 MALAYALAM LETTER OU 19 MALAYALAM LETTER AU 19 MALAYALAM LETTER AU 10 MALAYALAM LETTER AU	
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 O 14 MALAYALAM LETTER OU 15 MALAYALAM LETTER AU 16 MALAYALAM LETTER AU 17 MALAYALAM LETTER AU 18 MALAYALAM LETTER AU 19 MALAYALAM LETTER AU 19 MALAYALAM LETTER KA	
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 O 14 MALAYALAM LETTER O 15 MALAYALAM LETTER OU 16 MALAYALAM LETTER OU 17 MALAYALAM LETTER OU 18 MALAYALAM LETTER AU 19 MALAYALAM LETTER AU 19 MALAYALAM LETTER KA	
OB MALAYALAM LETTER VOCALIC R OC MALAYALAM LETTER VOCALIC L (This position shall not be used) OE MALAYALAM LETTER E OF MALAYALAM LETTER EE 10 MALAYALAM LETTER AI 11 (This position shall not be used) 12 MALAYALAM LETTER O 13 MALAYALAM LETTER O 14 MALAYALAM LETTER AU 15 MALAYALAM LETTER AU	
OD (This position shall not be used) OE MALAYALAM LETTER E OF 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 OU 15 MALAYALAM LETTER AU 15 MALAYALAM LETTER KA	
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	
11 (This position shall not be used) 12 MALAYALAM LETTER O 13 MALAYALAM LETTER OO 14 MALAYALAM LETTER AU 15 MALAYALAM LETTER KA	
12 MALAYALAM LETTER O 13 MALAYALAM LETTER OO 14 MALAYALAM LETTER AU 15 MALAYALAM LETTER KA	
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	
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) MALAYALAM LETTER VOCALIC RR MALAYALAM VOWEL SIGN VOCALIC LL MALAYALAM VOWEL SIGN VOCALIC LL (This position shall not be used) MALAYALAM DIGIT TONE MALAYALAM DIGIT TONE MALAYALAM DIGIT THREE MALAYALAM DIGIT FIVE MALAYALAM DIGIT FIVE MALAYALAM DIGIT FIVE MALAYALAM DIGIT SIEVEN MALAYALAM LETTER CHILLUN MALAYALAM LETTER CHILUN MALAYALAM

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

Ves

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?

Ves

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?

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?