# Proposal to encode Malayalam Consonant Sign Cillu

## 1.0 Introduction

The encoding of Malayalam Cillaksharams in Unicode is taking years for completion. The UTC decision to encode five Cillu letters with individual code points needs to be considered along with the available data on Cillu letters in Malayalam and its closely related Tamil Grantha scripts. Glyph shapes of many Cillu letters in the Malayalam and Tamil Grantha scripts are provided in a proposal for the first time. Tamil Grantha script used for millennia uses several Cillu letters to write Sanskrit texts, and Malayalam script also needs these Cillu letters to represent Sanskrit texts properly. Commonly, we find mention of only Malayalam texts in Malayalam script while discussing Cillaksharams in Unicode e-lists, but Malayalam has large collections of Sanskrit texts in manuscripts and books as well. The name, Cillu, instead of Chillu, is preferred and used in Unicode character names and in this proposal as the original Dravidian word, common to both Tamil and Malayalam, is written with the consonant, C and not with the aspirated consonant, Ch in Malayalam script.

It is requested that a MALAYALAM CONSONANT SIGN CILLU is encoded at U+0D4E in Unicode. This Malayalam Cillu Sign will be highly productive to add Cillaksharams from Malayalam or Sanskrit texts written in Malayalam script. If the individual Cillaksharam letters (U+0D7A .. U+0D7F) in the Unicode pipeline are also to be retained, a decomposition of the Cillaksharams using Cillu sign (U+0D4E) is also requested. A recent parallel example is that of the eleven Balinese letters which are given their canonical decompositions using a combining Tedung sign (see Appendix A) in the next upcoming version of Unicode 5.0 (Reference 2).

## 2.0 Malayalam Consonant Sign Cillu at U+0D4E

The 1970s introduction of letter reforms in Kerala state, India has produced different combinations of Cillaksharams and Candrakala pure consonants. As Prof. J. Richard Freeman, a Malayalam expert at University of Michigan wrote recently about the interchange of usage between pure consonants with virama and Cillaksharams: "They are just variable matters of convenience in the history of spelling words in the language, which is still changing."

K. P. Mohanan, Malayalam writing, (pp. 422-424, Daniels/Bright, World's writing systems, Oxford UP, 1996) "In the 1970s and 1980s, the difficulties of printing Malayalam script gave rise to the introduction of a simplified script. The most important property of the modern script is the linearization of the diacritics in such a way that a complex character can be built by a left-to-right sequence of separate sorts for the main symbol and the diacritics.
The second innovation is the breaking up of consonant clusters into sequences of atomic characters, using either a cillaksaram as in 'n' for 'nma', or the diacritic candra-kala (which otherwise writes /a/) to indicate a consonant without a vowel. Given these two changes, one would expect the modern script to become increasingly alphabetic, with each symbol representing a single segment. However, what has happened is that

individual printers have opted for "modernizing" some characters but not others, thereby creating an inconsistent script with a large number of random options."

Given this fluid situation between Cillaksharams and pure consonants (i.e. with Virama) in Malayalam usage, Malayalam linguists like Chitarajakumar (Rachana Aksharavedi group) and others have stated that the relationship between the base consonants and their corresponding Cillaksharams is explicitly maintained in Malayalam encoding. MALAYALAM CONSONANT SIGN CILLU (U+0D4E ) with canonical decompositions for individual Cillaksharams accomplishes this linguistic requirement. Cillaksharams are now created using ZW(N)J joiners in Unicode, and backward compatibility is supported to an extent with canonical decomposition of Cillakasharams using the combining Cillu sign. Another alternate option is to just encode the proposed MALAYALAM CONSONANT SIGN CILLU (U+0D4E) and define all the Cillaksharams in Named Sequences list. A parallel example is Haaru aksharams in Saurashtra script proposed in the Named Sequences list using SAURASHTRA CONSONANT SIGN HAARU (L2/06-251). The location of U+0D4E just one row below the Malayalam virama sign is chosen because in literature, the Cillu behavior is called "soft-virama".

Table 1. Malayalam Consonant Sign Cillu

| Unicode Character Name | Glyph Shape | Unicode Code Point |
|---|---|---|
| MALAYALAM CONSONANT SIGN CILLU | | U+0D4E |

## 3.0 Cillaksharam Letters in Malayalam and Tamil Grantha scripts

Tables 2 and 3 provide 10 Cillaksharams used in Malayalam script for Malayalam language texts. Note the "vertical stroke" towards the right end of the Cillaksharam letters, y, k, n, nn, l, ll, lll, r, t and m. The Cillu on letter "ma" (Table 3) is different graphically from the anuswara.

Table 2. Malayalam Cillaksharams
(H. Gundert, Reference 4, pg. xiii)

Cillu Y

TRANSLITERATION.                                      XIII

b. Final.              Cillu R

ക k    ൺ ṇ    ൻ ñ    ം=ം m    �യ y   ർ r    ൽ l    ൾ ḷ    ഴ l"

c. Followed by a Consonant.

ക്ത kta    ക്വ kva    ക്ഷ kṣa    ഗ്ന gna    ഗ്ധ gdha
ച്യ cya    ഛ്ര chra    ദ്ധ ddha    ജ്ഞ jña    ജ്ഝ jjha
ത്ഥ ttha    ത്ഥ്യ tthya    ത്ന tna    ദ്ധ ddha    ധ്ന dhna

2

# Table 3. Malayalam Cillaksharams
## (R. Grunendahl, Reference 5, pg. 92)

| k | ക് & ക് & ക് | r | ര് & ര് |
|---|---|---|---|
| ṇ | ണ് & ണ് | l | ല് & ല് |
| ṭ | ട് & ട് & ട് | ḷ | ള് & ള് |
| n | ന് & ന് | ḻ | ഴ് |
| m | മ് | | |

Table 4 provides the 22 Cillaksharams used in Tamil Grantha script for Sanskrit language texts. Note the "vertical stroke" in the Cillaksharams also. The 22 Cillaksharam consonants: k, g, ng, c, j, ny, tt, dd, nn, t, d n, p, b, m, y, r, v, sh, ss, s, and h.

## Table 4. Tamil Grantha Cillaksharams
## (R. Grunendahl, Reference 5, pg. 16)

| k | க் | n | ந் & ந் |
|---|---|---|---|
| g | கு & கு & கு | p | ப் |
| ṅ | ங் & ங் | b | ப் |
| c | ச் & ō | m | ம் |
| j | ஜ் | y | ய் |
| ñ | ஞ் & ஞ் | r | ர் |
| ṭ | ட் | v | வ் |
| ḍ | ட் & ட் | ś | ஶ் |
| ṇ | ண் | ṣ | ஷ் |
| t | த் | s | ஸ் |
| d | த் | h | ஹ் |

For the inter-transliterability between Malayalam and its closely related Tamil Grantha scripts, Malayalam Consonant Sign Cillu will be very useful. Tamil Grantha script is not yet encoded in Unicode and in its proposal, there will be a request seeking a parallel Tamil Grantha Consonant Sign Cillu. Sanskrit texts written in Malayalam script contain more Cillaksharams than what is sufficient for Malayalam texts as is clear from the comparison with Tamil Grantha cillaksharams.

If the six Cillaksaharams are also to be retained with individual code points in Unicode, it is recommended that they are provided with canonical decompositions using the proposed Malayalam consonant sign Cillu as shown in Table 5.

Table 5. Cillaksharam Decompositions

| Unicode code point for Cillaksharams. | Canonical decomposition with Malayalam consonant sign Cillu (U+0D4E) |
|---|---|
| U+0D7A (NN) | <U+0D23, U+0D4E> |
| U+0D7B (N) | <U+0D28, U+0D4E> |
| U+0D7C (R) | <U+0D30, U+0D4E> |
| U+0D7D (L) | <U+0D32, U+0D4E> |
| U+0D7E (LL) | <U+0D33, U+0D4E> |
| U+0D7F (K) | <U+0D15, U+0D4E> |

## 4.0 Summary

In sum, it is proposed that a Malayalam consonant sign Cillu (U+0D4E) be added which will help in handling large number of Cillaksharams needed to write Sanskrit texts in Malayalam script, and also in transliterating the many Cillaksharams available in the closely related Tamil Grantha script.

Table 5 provides canonical decompositions for Malayalam Cillaksharams in Unicode pipeline with the proposed combining Cillu sign in canonical equivalence as done for 11 Balinese letters (Reference 3). Another option is to generate all Malayalam Cillaksharams using Named Sequences definitions (Reference 2).

## 5.0 References

(1) E Muller, Malayalam chillaksharam, L2/06-207, 2006-05-16.
(2) N. Ganesan, Saurashtran Haaru consonants in Named Sequences list,
    L2-06/251, July 2006.
(3) Balinese script code chart and decomposition of 11 of its characters from,
    http://www.unicode.org/Public/5.0.0/charts/CodeCharts-5.0.0d2.pdf
(4) H. Gundert, Malayalam-English dictionary, Asian Educational Services reprint.
(5) Reinhold Grunendahl, South Indian Scripts in Sanskrit manuscripts and prints:
    Grantha Tamil - Malayalam - Telugu - Kannada - Nandinagari.
    (2001: Harrassowitz Verlag, Wiesbaden, Germany).

## Appendix A: Canonical Decompositions of 11 Balinese Letters

The new canonical decompositions for 11 of pre-composed Balinese
characters in Unicode 5.0 using Tedung sign (1B35) are:
1B06;BALINESE LETTER AKARA TEDUNG = <1B05 1B35>
1B08;BALINESE LETTER IKARA TEDUNG = <1B07 1B35>
1B0A;BALINESE LETTER UKARA TEDUNG = <1B09 1B35>
1B0C;BALINESE LETTER RA REPA TEDUNG = <1B0B 1B35>
1B0E;BALINESE LETTER LA LENGA TEDUNG = <1B0D 1B35>
1B12;BALINESE LETTER OKARA TEDUNG = <1B11 1B35>
1B3B;BALINESE VOWEL SIGN RA REPA TEDUNG = <1B3A 1B35>
1B3D;BALINESE VOWEL SIGN LA LENGA TEDUNG = <1B3C 1B35>
1B40;BALINESE VOWEL SIGN TALING TEDUNG = <1B3E 1B35>
1B41;BALINESE VOWEL SIGN TALING REPA TEDUNG = <1B3F 1B35>
1B43;BALINESE VOWEL SIGN PEPET TEDUNG = <1B42 1B35>


Like Balinese script example, there is another example of a South Asian islands script
with pure vowels being made up of a base letter and combining matras. In the right-to-left
Dhivehi script of  Maldives, characteristics of Arabic as well as South Indian scripts are
found. Unlike ISCII-derived Indic scripts of Unicode, Dhivehi possesses 24 pure
consonants in its writing system. And, a vowel-maatra for vowel /a/ along with other
vowel matras  added to a Thaana/Dhivehi consonant generates vocalised consonants
series as in Indic scripts, e.g.,ka, kaa, ki, kii, ku, kuu, ke, kee ... On the other hand, there
is no vowel maatra for /a/ in Indic Unicode. Pure vowels for the Thaana alphabet can be
accounted for in the future using the Zero-sound consonant, Alifu plus the combining
vowel matras. Its pure consonants are encoded in Unicode,  but pure vowels are not
encoded in Unicode are shown with combining vowel matras upon Alifu in,
http://homepage.ntlworld.com/stone-catend/Tha02.pdf

Naga Ganesan, PhD
naa.ganesan@gmail.com