L2/14-015

Proposal to encode MALAYALAM SIGN VERTICAL BAR VIRAMA

Cibu Johny, cibu@google.com
Shiju Alex, shijualexonline@gmail.com
Sunil V S, vssun9@gmail.com

2014-Jan-10

Introduction

The Virama in Malayalam script is called *Candrakkala* in its explicitly visible form: ◌ൌ. It has got the following two main functions:

1. Represent the neutral vowel sound (ə) at the word endings. This sound is called *Saṃvr̥tōkāraṃ* and transliterated as *ŭ*.[1]
2. Remove the inherent vowel A from a Malayalam consonant letter to indicate the corresponding pure consonant.

Candrakkala was introduced in Malayalam script, by Rev. Dr. Hermann Gundert around 1847 to represent *Saṃvr̥tōkāraṃ*. Before that, *Saṃvr̥tōkāraṃ* was implicit at word endings with nominal consonant letters and often contextual as well. After around half a century, by around 1900, *Candrakkala* acquired its the second function to represent pure consonants without their inherent vowel A.

Long before the time *Candrakkala* was used to represent pure consonants, vertical bar virama was used to indicate the pure consonants. Its usage was limited to transliterating foreign words from Sanskrit, Tamil or European languages.

Its appearance is as a long vertical bar above the letter at its center. Older attestations allowed this vertical bar to cut across the consonant letter; however, later attestations do not connect or ligate with the consonant letter.

Attestations of Vertical Bar Virama range from 1700s to 1860s, while that of *Candrakkala* starts at around 1850. This symbol was almost out of use by the time *Candrakkala* was introduced. Once *Candrakkala* assumed its second functionality to represent pure consonants this sign became redundant as well.

Vertical bar virama must be encoded to record the documents from its above usage period.

[1] As per ISO 15919 described in http://homepage.ntlworld.com/stone-catend/trind.htm

Comparison to Candrakkala

Both Vertical Bar Virama and *Candrakkala* can indicate the removal of default vowel A from the the nominal consonant letters. However the differences are plenty. Vertical Bar Virama is never used to indicate *Saṃvṛtōkāraṃ* which was the original purpose of *Candrakkala*. Vertical Bar Virama is never found in native words; it is only seen in transliteration of words in Sanskrit, Tamil and European languages.

They both vary in shape and their position above the letter. While Vertical Bar Virama is always found at the top-middle position, *Candrakkala* is currently found only at the top-right corner of the letter. These two are from two different era as well. While the attestations for Vertical Bar Virama are between 1700-1868, we have good indications that, it was existing before that period as well. On contrast, *Candrakkala* was invented at around 1850 and its original purpose was to indicate *Saṃvṛtōkāraṃ*.

Comparison to Circular Virama

Circular Virama is proposed as a separate document.

Both Vertical Bar Virama and Circular Virama indicates the removal of default vowel A from the the nominal consonant letters. However there are some differences as well. So far we have not seen Vertical Bar Virama in native words; it is only seen in transliteration of other language words in Sanskrit, Tamil, European languages etc.. However, Dr. Hermann Gundert asserts that this sign was used in Northern Kerala. Circular Virama at the same time, is heavily used for native Malayalam texts. One important usage was to split up consonant clusters, probably to be reduce number of glyphs required for printing.

They both vary in shape and position relative to the letter. Vertical Bar Virama is always found above the letter. Vertical Bar Virama is positioned at the center of the letter while Circular Virama is placed at the top-right corner.

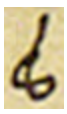
While the attestations for Vertical Bar Virama are between 1700-1868, we have good indications that, it was existing before that period as well. At the same time, Circular Virama appeared at around 1860 and disappeared by 1920 when *Candrakkala* assumed its functionality.

Relationship to Chillus

It can be imagined that Vertical Bar Virama is the precursor in the evolution of Chillus. Chillus can be thought as the ligatured form of a consonant letter with Vertical Bar Virama. However, some Chillus evolved significantly from there. Some of them are now associated with a base consonant different from its original base. For example, ligatured form of <TA, Vertical Bar Virama> became the CHILLU L:

 evolved into ൽ

Similarly, Chillu LL is derived from <TTA, Vertical Bar Virama>:

 evolved into ൾ

At the sametime, some other chillus just straightforward ligature with Vertical Bar Virama. Examples: ന → ൻ, ണ → ൺ.

Since Chillus are now atomically encoded, this poses a dilemma on the treatment required for Vertical Bar Virama. Vertical Bar Virama can be applied not only to single consonant letters, but also to consonant clusters like 'ന്ത'(N-TA) as in the last attestation below. So it is very inefficient to exhaustively encode every possible permutation of consonant clusters with Vertical Bar Virama. That means encoding Vertical Bar Virama is inevitable and that can be achieved by treating it differently from the Chillus that should be atomically encoded. The model proposed by Rev. Dr. Hermann Gundert seems reasonable:

In 1851 *Malayāḷābhāṣāvyākaraṇaṃ* he says:
"To indicate a half consonant without the vowel, elongate the tail for ക് (CHILLU K), ൎ (CHILLU NN), ൻ (CHILLU N),  *(CHILLU M),*  *(CHILLU Y), ർ (CHILLU RR), ൽ (CHILLU L), ൾ (CHILLU LL),* *(CHILLU LLL) and for rest of them add a combining mark as in*  *<TTA, Vertical Bar Virama> and*  *<PA, Vertical Bar Virama>."*

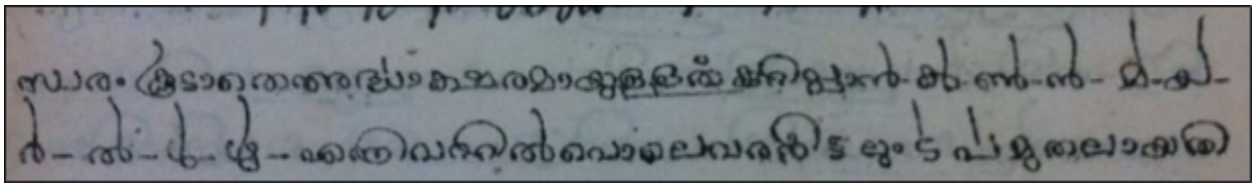


Image 1

Even though the earliest attestations of Vertical Bar Virama (in documents from Johann Hanxleden), has it stretching across the letter, later authors does not do that. They use it as a non-ligating combining mark. For them, for example,  (<sa, veritcal bar virama>) does not look like a Chillu.

Confusability with Dot Reph(0D4E)

It is true that in some occasions authors write Vertical Bar Virama that could be confusable with DOT REPH(0D4E). For example, consider the attestation from Benjamin Bailey's:

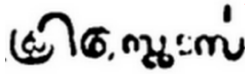


Image 2

The word is clearly 'Christos' from the context. Without the contextual information, reader could confuse it as 'Christorsa', by reading Vertical Bar Virama as DOT REPH.

To avoid such security issues, this character should be excluded from IDNs since it is not needed for contemporary Malayalam use.

Attestations

Attestations from various sources are listed below as sources organized as subsections.

Grammatica Grandonica[2] by Toon Van Hal & Christophe Vielle

This book reproduces the original manuscript by Johann Ernst Hanxleden S.J. (1681–1732).

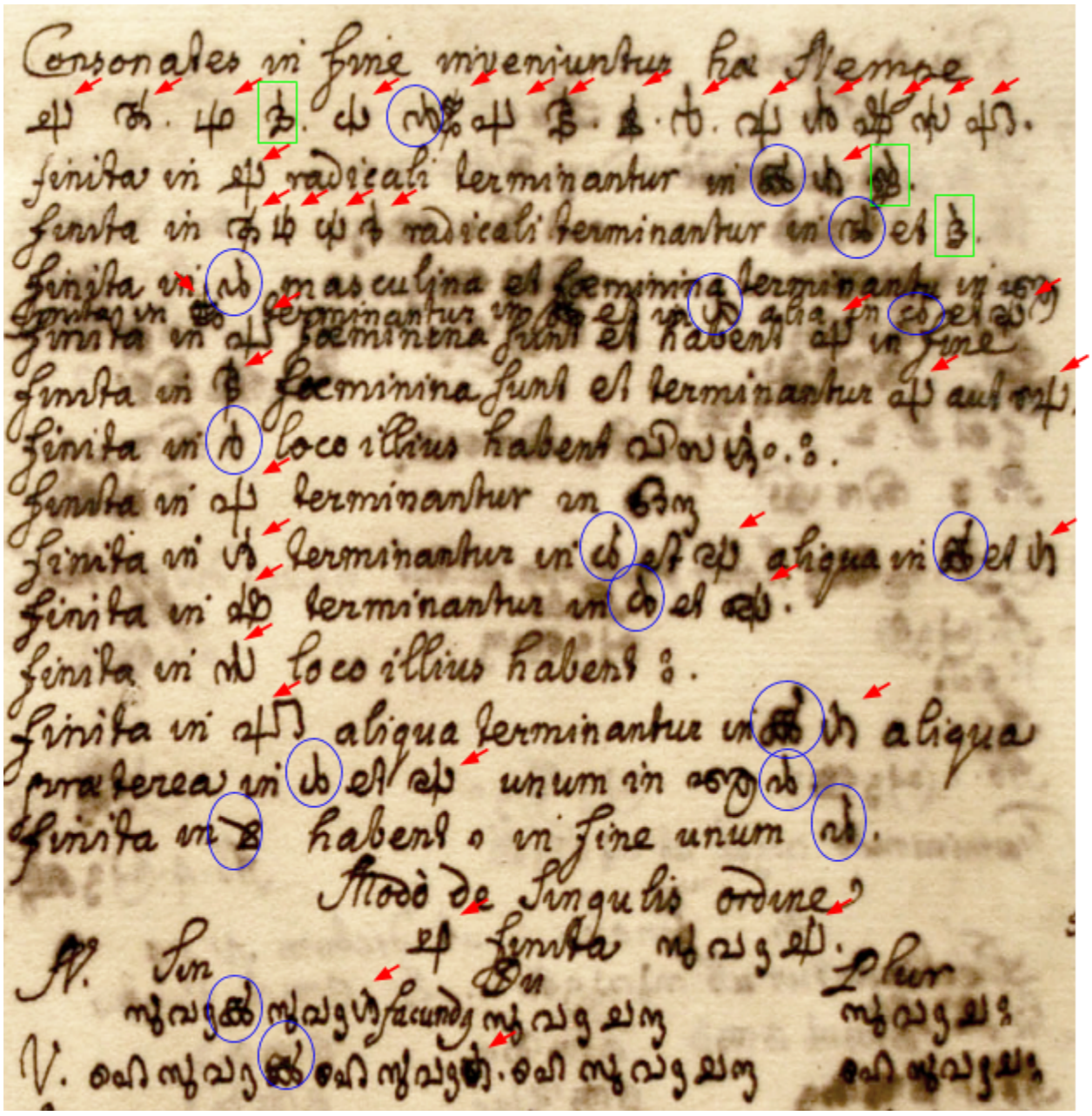


Image 3: Page 60 lists various word final vowelless forms possible in a Sanskrit. The consonant letters with Vertical Bar Virama are marked in red. They are: ഗ ച ജ ഡ ത ഥ ദ ധ പ ബ ഭ മ വ ശ ഷ സ ഹ. Ligatured forms that are yet to be proposed as chillus are marked in green box: ങ ദ. The existing or proposed chillus are marked in blue circle.

[2] http://opus.kobv.de/ubp/volltexte/2013/6321/pdf/hanxleden_grammatica.pdf

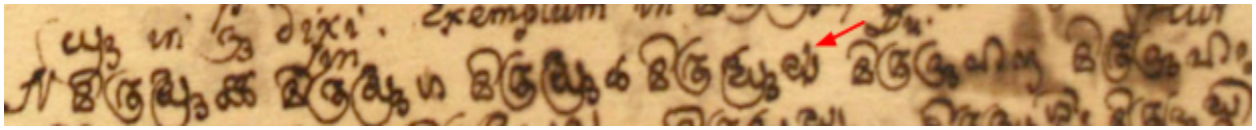


Image 4: Page 76 shows Vertical Bar Virama with ഡ(DA) in a word. This book has many such examples.

The authors Toon Van Hal and Christophe Vielle describe their struggle on reproducing the the manuscript using Indolipi software that uses Unicode:

> It was not possible to reproduce all ancient Grantha Malayāḷam characters. For avoiding confusion, we do not use the pure final *t* (written like ൽ in the manuscripts) and replace it by the modern script ൻ (= Mal. *tŭ*), because in Malayāḷam script the character ൽ is used for final *l* (cf. the confusion by Hanxleden himself in §9). Similarly in the case of final *ṭ* (written like ൾ in the manuscripts), replaced by the modern script ർ, because in Malayāḷam script the character ൾ is used for final *ḷ* (e.g. *-ādikaḷ* in §81).[32] In addition to the still-extant characters for final *-r* (ർ) and *-n* (ൻ), we can make use of a special one for final *-k* (ക്). Hanxleden, however, renders other final consonants by means of an oblique stroke within the consonantal sign (see e.g. on ff°9r, 41r-v). Such *virāma*'s are replaced by the sign ˘ following the final consonants.

Image 5: From page 18 of the book.

Encoding Vertical Bar Virama could help tremendously in endeavors like this book.

Ceṟupaitaṅṅaḷkka Upākārartham imklīśilninna Paribhāṣappetuttiya Kathakaḷ (1824), Benjamin Bailey[3]

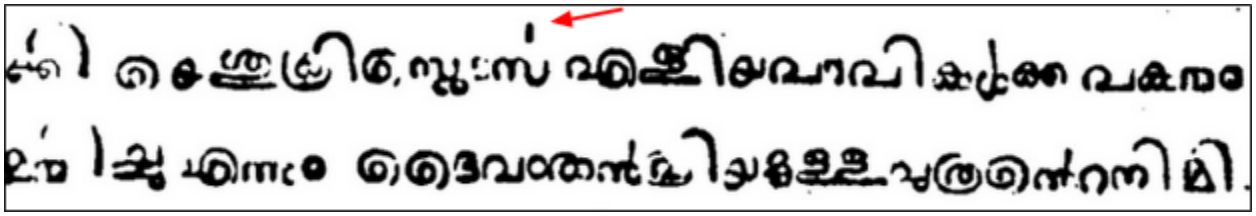


Image 6: It shows the usage of Vertical Bar Virama to transliterate the European name 'Christos'.

Malayāḻābhāṣāvyākaraṇaṃ (1851), Hermann Gundert

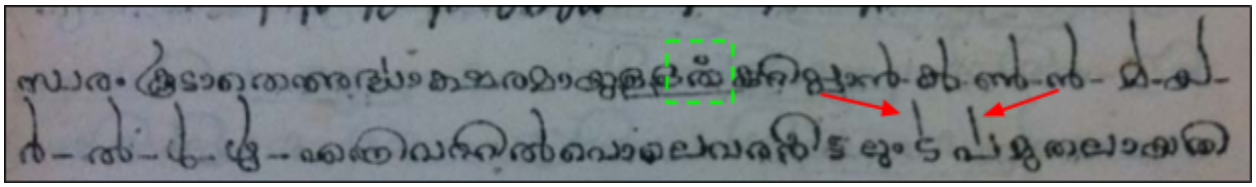


Image 7: From page 3 of the book. The contrasting usage of *Candrakkala* and Vertical Bar Virama. The Candrakkala in dotted box indicates *Saṃvṛtōkāraṃ*. Vowelless consonants of ട (TA) and പ (PA) are indicated

[3] http://www.benjaminbaileyfoundation.org/benjamin_books.php?id=17 Page 83 which reproduces the page 3 of the original document.

by Vertical Bar Viramas marked by red arrows.

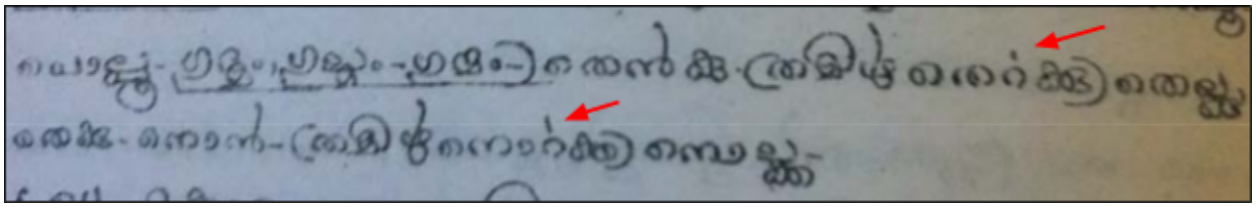


Image 8: From page 12 of the book. Vowelless consonants of ഠ (RRA) is indicated by Vertical Bar Viramas marked by red arrows. This is used in the transliteration of the Tamil word தெற்கு (*teṛkku*).

Grammar of Malayalam Language (1868), Hermann Gundert

This is the print version of *Malayāḻābhāṣāvyākaraṇaṃ* (1851), by Herman Gundert. The usage of Vertical Bar Virama is continued for the text in page 3:

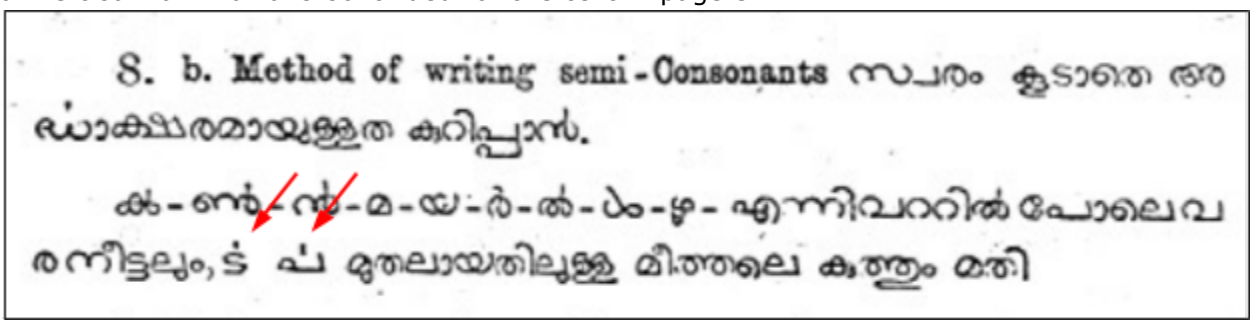
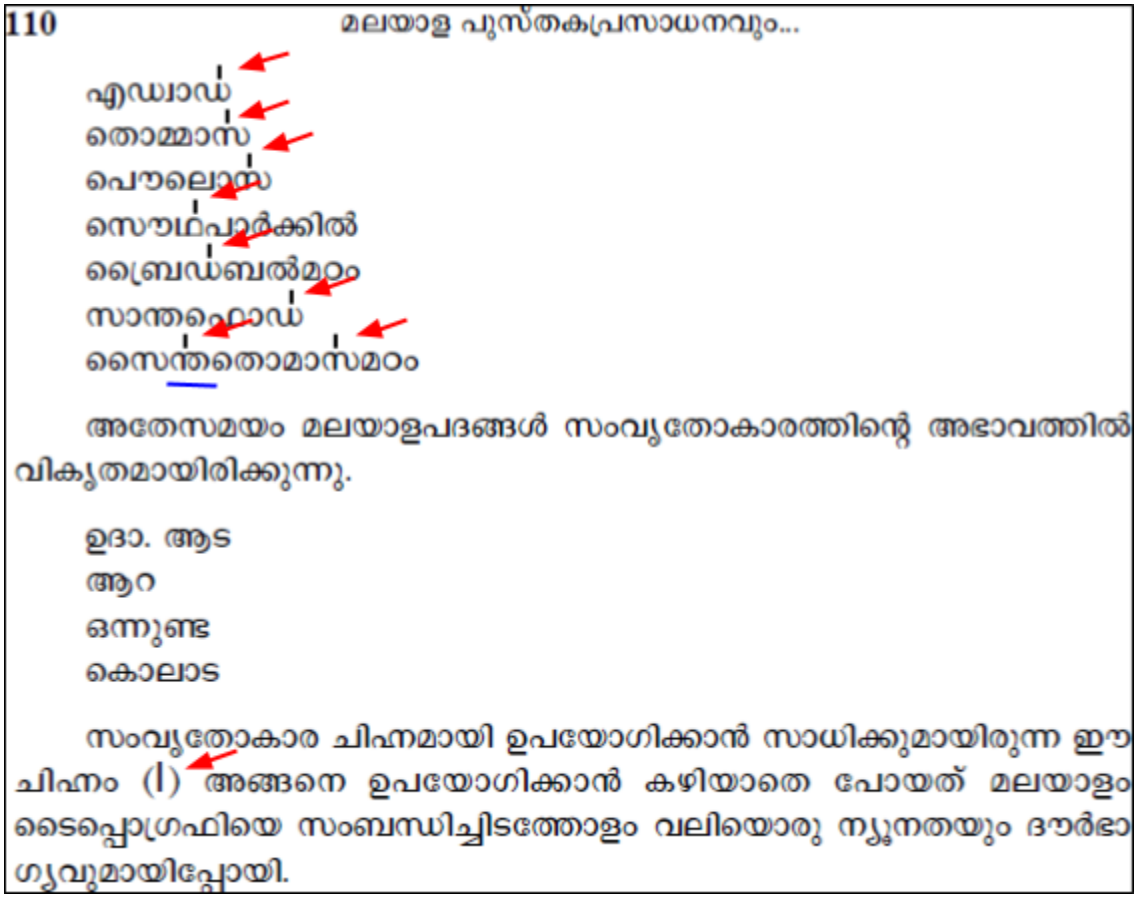


Image 9

The second attestation in page 12 of *Malayāḻābhāṣāvyākaraṇaṃ* is converted to *Candrakkala* in this print version.

Beṅcamin Beyili(2009), by Babu Cheriyan[4]


110 മലയാള പുസ്തകപ്രസാധനവും...

എഡ്വാഡ്
തൊമ്മാസ്
പൌലൊസ്
സൌഥ്പാർക്കിൽ
ബ്രൈഡ്ബൽമഠം
സാന്തൈഹാഡ്
സൈന്തതൊമാസ്മഠം

അതേസമയം മലയാളപദങ്ങൾ സംവ്യതോകാരത്തിന്റെ അഭാവത്തിൽ വികൃതമായിരിക്കുന്നു.

ഉദാ. ആട
ആറ
ഒന്നുണ്ട
കൊലാട

സംവ്യതോകാര ചിഹ്നമായി ഉപയോഗിക്കാൻ സാധിക്കുമായിരുന്ന ഈ ചിഹ്നം (l) അങ്ങനെ ഉപയോഗിക്കാൻ കഴിയാതെ പോയത് മലയാളം ടൈപ്പൊഗ്രഫിയെ സംബന്ധിച്ചിടത്തോളം വലിയൊരു ന്യൂനതയും ദൗർഭാഗ്യവുമായിപ്പോയി.

Image 10: Use of Vertical Bar Virama in this book is to illustrate its archaic usage by Benjamin Bailey in 'Ceṛupaitaṅṅaḷkka...'. Please note that it is used on top of consonant clusters as well (underlined). This document also shows the need to encode this character to accurately record the texts from that period.

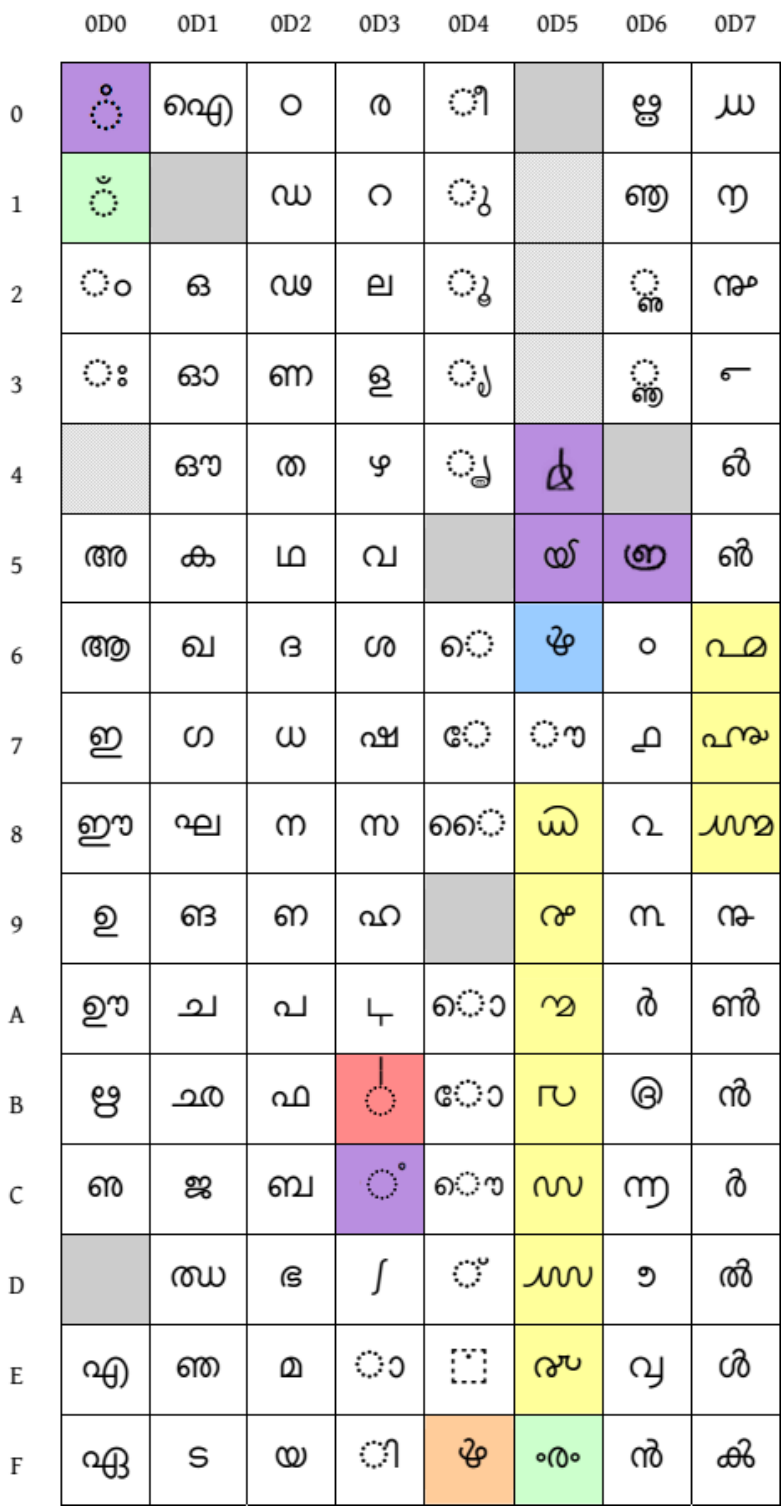
Based on the above attestations we can see that this character was in use till around 1868. Since then it has been replaced by the Virama: ◌് (0D4D).

[4] http://www.benjaminbaileyfoundation.org/benjamin_books.php?id=17 page 110

Choice of codepoint

The proposed codepoint is **0D3B** which is close to the locations of Virama (0D4D) and the proposed Circular Virama(0D4C). It is also close to the proposed Circular Virama (0D3C) location.

Malayalam Code Chart



Legend

White: Published as of Unicode 6.3.

Green: Accepted and at advanced stage not permitting changes.

Yellow: Fractions, accepted.

Orange: Chillu LLL accepted but preferably to be moved to group with Chillu Y.

Blue: New location proposed for Chillu LLL.

Purple: Locations of ongoing proposals of Anusvara Above, Chillus M and Y and Circular Virama and Symbol Para.

Red: Location proposed for Bar virama.

Proposed character with properties

0D3B;MALAYALAM SIGN VERTICAL BAR VIRAMA;Mc;0;L;;;;;N;;;;;

Official Proposal Summary Form

(Based on N3902-F)

A. Administrative

1. Title

Proposal to encode MALAYALAM SIGN VERTICAL BAR VIRAMA

2. Requester's name

Cibu Johny, Shiju Alex, Sunil V S

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

Individual contribution

4. Submission date

2014-Jan-10

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

Yes, Malayalam

2. Number of characters in proposal

1 (one)

3. Proposed category

Category B1, specialized small (for this character)

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?

Michael Everson

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

Michael Everson, everson@evertype.com

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.

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?

The authors themselves are part of the user community.

2c. If YES, available relevant documents

None specifically. The matter was discussed via email.

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

Those (scholars and enthusiasts) who desire to store old Malayalam documents using this character.

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

Rare

4b. Reference

See detailed proposal.

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

No. This character was in historical use.

5b. If YES, where?

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?

It belongs in the Malayalam block which is in the BMP.

6c. If YES, reference

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

Only one character is proposed.

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?

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?
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?
13c. If YES, reference: