Proposal to encode Malayalam minor fractions

Shriramana Sharma, jamadagni-at-gmail-dot-com, India

2013-Apr-25

L2/13-051R

§1. Thanks

I thank Anshuman Pandey for drawing my attention to Malayalam fractions, Cibu Johny for providing attestation samples, and Elmar Kniprath for background checks into dictionaries.

§2. Background

Recently I submitted a proposal to encode Tamil fractions and symbols (see L2/12-231 and L2/13-047). I had noticed that Malayalam already has the major fractions ¼, ½ and ¾ encoded (based on which I had submitted L2/09-376 for the Tamil major fractions alone). I later found out that Malayalam also has a rich set of minor fractions while examining other earlier documents. Despite their introduction as far back as L2/05-152, they have not yet been encoded. I hence submit this proposal to encode 10 minor fractions for Malayalam.

It is to be noted that these same fractions (as mentioned in L2/05-152) are also shown in the Kerala Govt's documents L2/05-175 and L2/05-249, of which only the major fractions were formally proposed in the summary document L2/05-295 and subsequently encoded. The present proposal seeks to complete the set of Malayalam fraction characters.

§3. List of characters

The Tamil/Grantha/Malayalam region has historically always been very close in various matters, of which script is one. It is hence not surprising that Malayalam should also contain the same set of minor fractions as in Tamil and with the same names (with of course slight morphological changes).

The list of Malayalam fractions is as follows:

	Glyph	x/320	Value	Words	Malayalam name
1.	ൣ	1/320	1/320	one three-hundred-and-twentieth	muntiri
2.	ൡ	2/320	1/160	one one-hundred-and-sixtieth	arakkāṇi
3.	ൠ	4/320	1/80	one eightieth	kāṇi
4.	ൟ	8/320	1/40	one fortieth	aramā

5.	ൗ	12/320	3/80	three eightieths	mūnnukāṇi
6.	ൊ	16/320	1/20	one twentieth	orumā
7.	ോ	20/320	1/16	one sixteenth	mākāṇi
8.	ൎ	32/320	1/10	one tenth	raṇṭumā
9.	൏	40/320	1/8	one eighth	arakkāl
10.	ൔ	48/320	3/20	three twentieths	mūnnumā
11.	ൕ	60/320	3/16	three sixteenths	muṇṭāṇi
12.	൐	64/320	1/5	one fifth	nālŭmā
13.	ൾ	80/320	1/4	one quarter	kāl
14.	ൽ	160/320	1/2	one half	ara
15.	ൿ	240/320	3/4	three quarters	mukkāl

A comparison with the list of Tamil fractions (L2/12-231 p 3) indicates that the only missing items are kālvīcam $^{1}/_{64}$, araivīcam $^{1}/_{32}$ and mukkālvīcam $^{3}/_{64}$. As we have already noted (ibid p 8), even in Tamil these “were probably a later development for convenience”. Therefore it is possible that they were not in use in Malayalam.

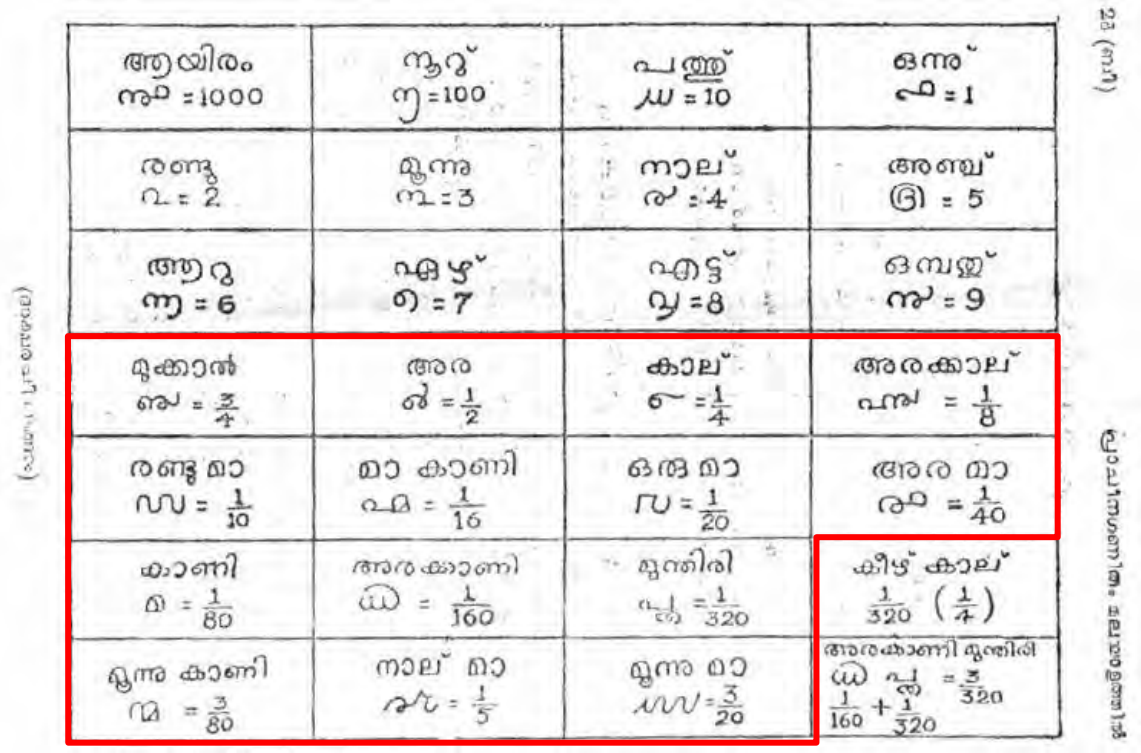
Of the above written forms, the last three, the major fractions, are already encoded at 0D73, 0D74 and 0D75. The sign for kāṇi $^{1}/_{80}$ is merely 0D2E LETTER MA മ (or a slight glyphic variant thereof). The sign for the smallest fraction muntiri $^{1}/_{320}$ is merely a stacked presentation of the syllable P·TA (0D2A 0D4D 0D24): പ്ത

Note that in current Malayalam orthography this stack may not be used and the cluster written as പ്ത. Hence fonts may not provide the stacked presentation by default. In this case, the accepted sequence used to request C2-conjoining forms may be used to request the stack: PA + ZWJ + VIRAMA + TA (see for instance TUS 6.1 p 299 for Bengali R·YA or p 315 for Telugu R·MA). Of course, this requires the font to support this sequence.

As per the guidance received from the UTC regarding Tamil letters which are additionally used for fractions or symbols (see L2/13-047 §1.1 p 1) I have not proposed these two written forms (i.e. P·TA for $^{1}/_{320}$ and MA for $^{1}/_{80}$) as separate characters. If the UTC deems that the situation here is different, I request to be informed so I can change this.

As such, with the current situation, 10 new characters are proposed to be encoded.

§4. Attestations



ആയിരം ൱ =1000	നൂറ് ൲ =100	പത്ത് ൰ = 10	ഒന്ന് ൯ = 1
രണ്ട് ൨ = 2	മൂന്ന് ൩ = 3	നാല് ൪ = 4	അഞ്ച് ൫ = 5
ആറ് ൬ = 6	ഏഴ് ൭ = 7	എട്ട് ൮ = 8	ഒമ്പത് ൯ = 9
മുക്കാൽ ൾ = $\frac{3}{4}$	അര ൹ = $\frac{1}{2}$	കാല് ർ = $\frac{1}{4}$	അരക്കാല് ൻ = $\frac{1}{8}$
രണ്ടു മാ ൘ = $\frac{1}{10}$	മാ കാണി ൙ = $\frac{1}{16}$	ഒരു മാ ൗ = $\frac{1}{20}$	അര മാ ൚ = $\frac{1}{40}$
കാണി ൛ = $\frac{1}{80}$	അരക്കാണി ൜ = $\frac{1}{160}$	മുന്തിരി ൝ = $\frac{1}{320}$	കീഴ് കാല് $\frac{1}{320}$ $(\frac{1}{4})$
മൂന്നു കാണി ൞ = $\frac{3}{80}$	നാല് മാ ൟ = $\frac{1}{5}$	മൂന്നു മാ ൠ = $\frac{3}{20}$	അരക്കാണി മുന്തിരി ൜ ൝ = $\frac{3}{320}$ $\frac{1}{160} + \frac{1}{320}$

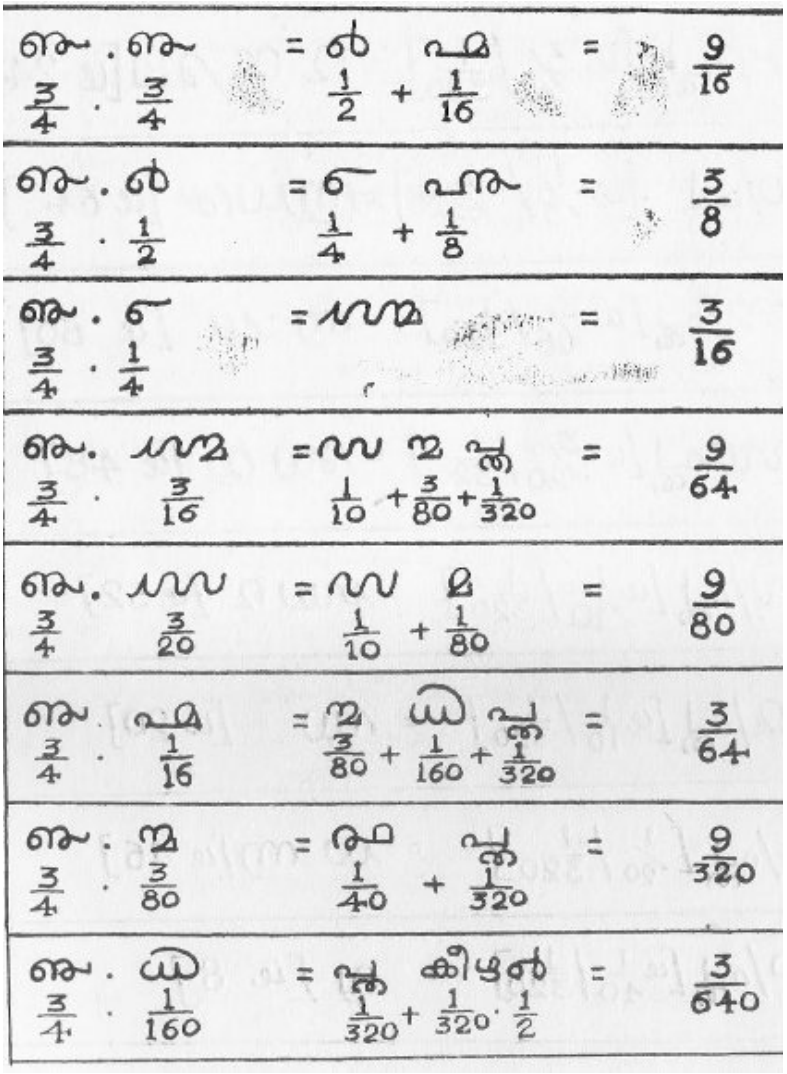
Figure 28b,

Prācīna Gaṇitaṃ Malayāḷattil, Prof C K Moosathu, Kerala State Institute of Language, 1980.

(cited in Kerala State Govt docs L2/05-175 p 3 and L2/05-249 p 4,

also in Peter Constable's L2/05-152 p 2, UTC's L2/05-295 p 3)

Note that this does not show ൠ൨ $^{3}/_{16}$. It shows all other fractions with their names.

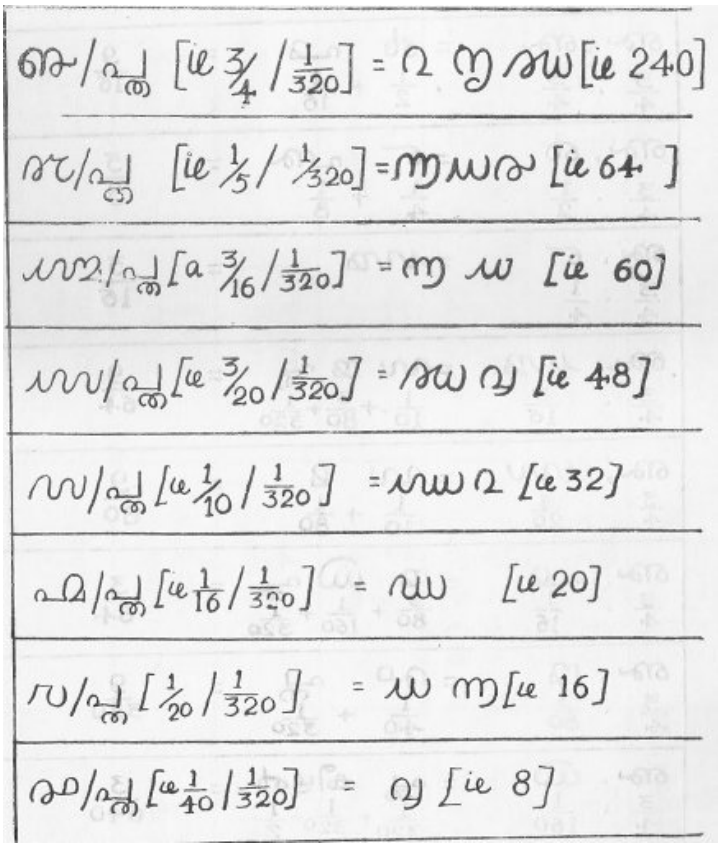


Ibid fig 28c.

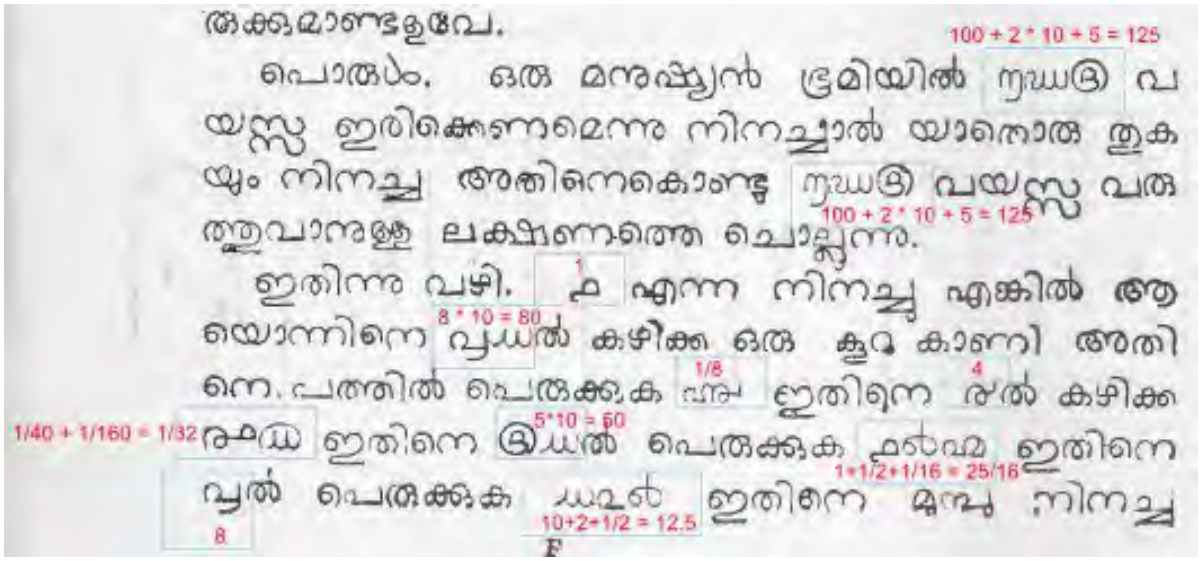
This sample shows ൗ ³/₁₆ (third section from top). This fraction is known as muṇṭāṇi, as can be verified from sources like Hermann Gundert's Malayalam dictionary of 1872.

This sample also illustrates actual usage of the fractions.

Further, this sample (bottom-most section, before ൽ ¹/₂) also shows the word kīḻ (spelt as kīḻa for some reason) used to downscale the fraction by the magnitude of the lowest minor fraction: ¹/₃₂₀. The system used is apparently identical to that in Tamil and more details of the same may be obtained from my Tamil fractions proposal L2/12-231 §4.5 p 10. However unlike in Tamil so far no attestation has been obtained in Malayalam for a separate abbreviated symbol for the downscaling term kīḻ.



Ibid fig 28c.



Kaṇakkadhikāram, Mānavan Māppilla,

cited in aforementioned Kerala Govt docs

§5. Characters *not* proposed

§5.1. Pending further research

Yet another source "Kēraḷattilē prācīna lipi mātṛkakaḷ" by one Dr N Sam, 2006, seems to show (on p 72) some additional glyphs for some of the lower fractions: kīḻ muntiri ($^1/_{320}$ × $^1/_{320}$), kīḻ ara(y)kkāṇi ($^1/_{320}$ × $^1/_{160}$), kīḻ oruma ($^1/_{320}$ × $^1/_{80}$), kīḻ araykkāl ($^1/_{320}$ × $^1/_8$), kīḻ kāl ($^1/_{320}$ × $^1/_4$), kīḻ ara ($^1/_{320}$ × $^1/_2$) and kīḻ mukkāl ($^1/_{320}$ × $^3/_4$) – these are marked below in red:

ഭിന്നസംഖ്യകൾ			
കീഴ്മുന്തിരി	അരമ (1/40)	മാകാണി (1/16)	മുക്കാൽമാകാണി
കീഴ്അരയ്ക്കാണി	"	"	"
കീഴ്ഒരുമ	മുക്കാണി (3/80)	അരയ്ക്കാൽ	നാണയങ്ങൾ
കീഴ്അരയ്ക്കാൽ	ഒരുമ (1/20)	"	രൂപ
കീഴ്കാൽ	"	കാല് (1/4)	പണം
കീഴ്അര	ഇരുമ (1/10)	അര (1/2)	" (കൊച്ചി)
കീഴ്മുക്കാൽ	"	"	" (കൊച്ചി)
മുന്തിരി (1/320)	മൂന്നുമ (3/120)	മുക്കാൽ (3/4)	ചക്രം
അരയ്ക്കാണി (1/160)	മുണ്ടാണി (3/16)	"	കാശ്
കാണി (1/80)	നാന്മ (1/5)	"	

However many of the glyphs in the above chart are highly similar if not identical to Tamil written forms and fractions. For instance $^1/_{20}$ oruma, $^1/_{10}$ irumā and $^3/_{20}$ mūnnumā are written quite similar to their Tamil forms ௳, ௴ and ௵. As such, further research is required to ascertain to which script these characters belong so that they may be appropriately encoded. The above table also contains some glyphs for currency (marked in blue) whose status is the same.

The same source also shows (on p 73) other glyphs for measures. (Compare the Tamil measures list under L2/12-231 §3.2 p 4.) However some of these are not sufficiently legible and some of which are again highly similar to Tamil/Grantha numerals and symbols and hence require the same verification as for the lower fractions as mentioned before:

ദ്രാവക/ധാന്യ അളവുകൾ		
ഉരി	കുറുണി	ചകട്ട
"	"	കലം
ആഴക്ക്	പതക്ക്	"
ഉഴക്ക്	ഇരുകുറുണി	"
"	മുക്കുറുണി	"
മുഴക്ക്	നാങ്കുറുണി	തുണി
"	അയ്ങ്കുറുണി	"
നാഴി	അറുകുറുണി	ഇരുതുണി
"	പടി	

One however notes that the names of the measures are again highly similar to Tamil: uri, āl̲akkŭ, ul̲akkŭ, mūl̲akkŭ, nāl̲i, kuruṇi, patakkŭ, ... paṭi, ... kalam, tūṇi etc. Nevertheless, the script identity of the glyphs must be verified. It would also be preferable to ascertain the value of these measures (again, comparing with Tamil) before they are encoded.

§5.2. Not valid candidates for encoding

The sources also attest glyphs for the numerals 20, 30, 40, ... 90 (ibid p 65):

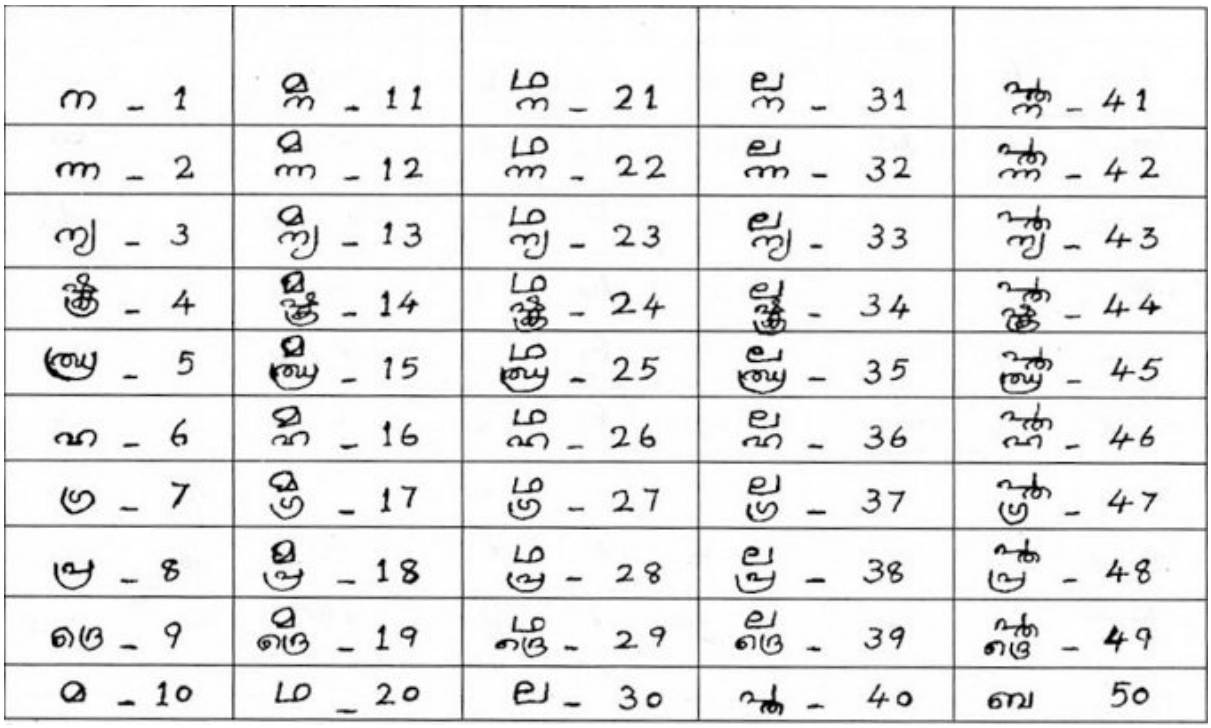
However, a close examination will reveal that these are merely ligatures of 2 ൨, 3 ൩, 4 ൪ etc with ൰ for 10. For instance for twenty we fuse 2 and 10 i.e. ൨ + ൰ to get ൨൰.

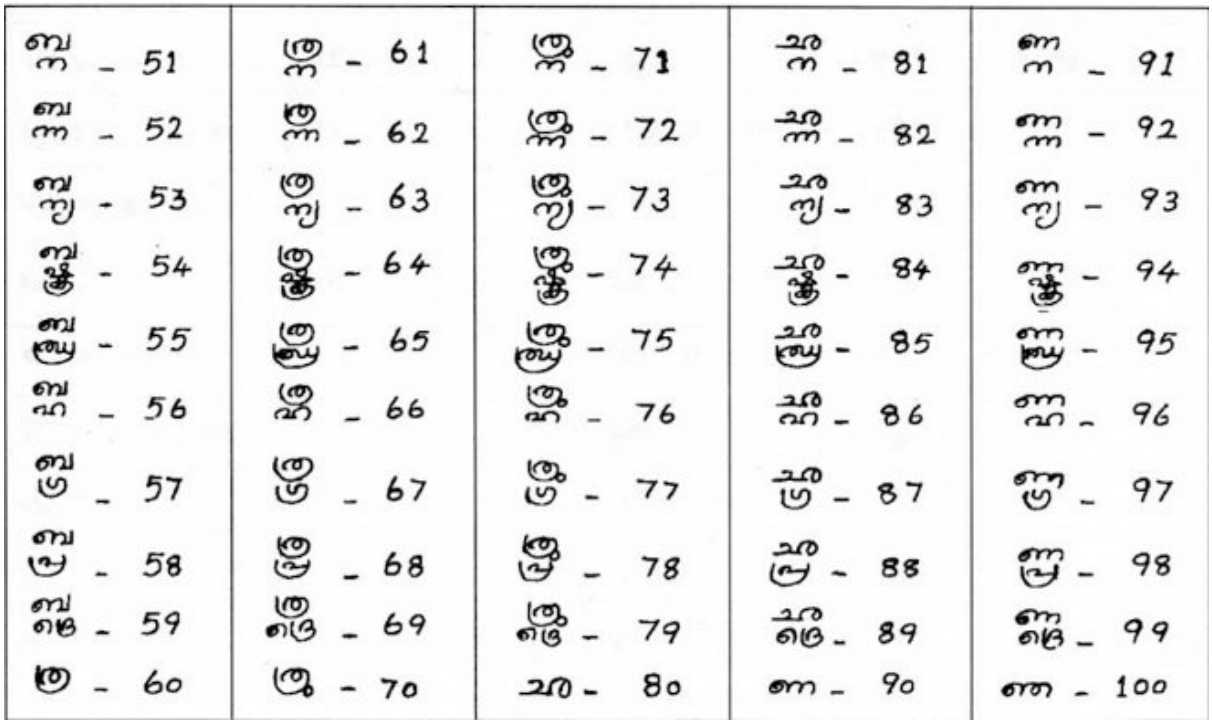
As such these are not valid candidates for separate encoding. An appropriate font catering to this old orthography may merely substitute the sequences 0D68 TWO + 0D70 TEN etc with the appropriate ligated glyph ൨൰ etc.

It would be useful to document this fact in the Unicode chapter on Malayalam.

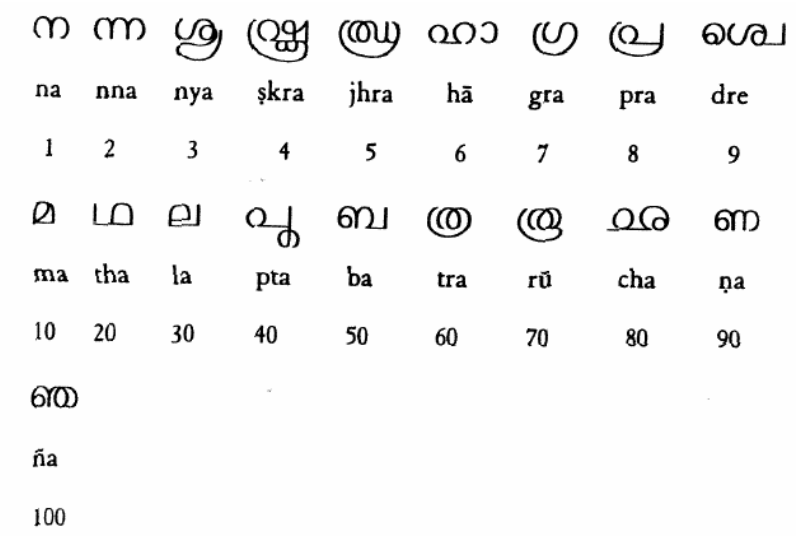
There is also an old system of denoting numerals entirely by letter forms called akṣarapallī. This is unlike the prevalently used system called aṅkapallī using digits where at best some digits are just similar to letters. In akṣarapallī, outright letter forms are used for numbers.

Ibid pp 70,71 gives the numbers from 1 to 100 in this system:





These were also mentioned in L2/05-173 and L2/06-260. The latter also provides samples (which I separately verified via Elmar Kniprath) from Reinhold Gruenendahl's 2001 work "South Indian Scripts in Sanskrit Manuscripts and Prints" (ISBN 3-447-04504-3) p 94:



(The above sample gives the wrong Malayalam glyphs for nya 3 and dre 9, but the transliterations correctly correspond to the numerical values.)

Based on this attestation, L2/06-260 had suggested that these characters be encoded. However, they are clearly a system of re-use of letters for numbers and cannot be encoded. Even though the previous samples show non-standard ligating and stacking behaviour (see for example the lower element of the stack denoting 9 in 19, 29 etc taking a vowel sign), each written form is unambiguously associated with a numerical value irrespective of its position and presumably a linear presentation would also not be unacceptable. In any event, higher level protocols than encoding can take care of this stacked presentation and it is not necessary for the encoding model to cater to it.

It is to be noted that this akṣarapallī system was evidently devised by scholars who needed to include numbers into verses. For instance, mathematical or astronomical treatises would require an easy way of representing numbers (rather than by their full words), and the fact that a great amount of Sanskrit scientific literature is in the form of metered verse would require that letters be used and also restrict their count. Thus the letter forms (whether simple letters or ligatures) which were most similar to the separate written forms of the numbers were used to denote those numbers in the olden days. (Ref:

Appendix on Palaeography in “South Indian Temple Inscriptions”, T N Subramanian, Government Oriental Manuscripts Library, Madras, 1953-1957, p 1567.) Thus, the specific letter forms came to be associated with the specific numbers and this notation continued to be used even after the shapes of the letter forms and those of the digits diverged in their separate evolutions. As a result, not only in Malayalam but old Indic manuscripts in other Indic scripts exhibit this system of denoting numbers too.

Therefore this is to be merely seen as a re-use of letter forms and hence does not warrant separate encoding.

§6. Nature of proposed encoding

As mentioned at the outset, this document proposes to encode only the 10 minor fractions which are clearly attested and not representable by existing letter characters. There are more than ten empty codepoints in the Malayalam block but not contiguously. Therefore it is not possible to encode all the proposed characters contiguously. Due to the availability of sufficient codepoints in the existing Malayalam block there is no need for a supplementary block unlike what was requested for Tamil in L2/12-231 (p 2). A judicious allotment of codepoints is found to be sufficiently meaningful and practicable.

The proposed allotment is as follows. (For reference see the chart on p 13.)

It is noted that the three already-encoded major fractions are in the eighth column of the Malayalam block at 0D73, 0D74, 0D75. There are three empty codepoints next to those at 0D76, 0D77, 0D78. In Oriya, the fractions $^{1}/_{16}$, $^{1}/_{8}$ and $^{3}/_{16}$ are found in addition to the major fractions $^{1}/_{4}$, $^{1}/_{2}$ and $^{3}/_{4}$ and they are encoded immediately after them (in that order) from 0B72 to 0B77. The same would make sense for the Malayalam fractions $^{1}/_{16}$, $^{1}/_{8}$ and $^{3}/_{16}$ as well (since in any case we cannot place all the minor fractions together).

The rest of the minor fractions are proposed to be placed in ascending order of value (as seen in §3 of this document) at the bottom of column 6 (i.e. 0D5x, excepting 0D5F allotted for ARCHAIC II in the pipeline) where there is sufficient space and which would also be visually close to the Malayalam digits. This encoding still leaves sufficient (contiguous) space in the Malayalam block for future additions as well if necessary.

Appropriate informative aliases, annotations and cross-references to existing characters (where applicable) should be added to the proposed characters. The existing fractions should also be provided such informative aliases giving their Malayalam names.

§7. Unicode Character Properties etc

(These may also be found as attachments to this PDF.)

§7.1. Additions to UnicodeData.txt

```
0D58;MALAYALAM FRACTION ONE ONE-HUNDRED-AND-SIXTIETH;No;0;L;;;;1/160;N;;;;;
0D59;MALAYALAM FRACTION ONE FORTIETH;No;0;L;;;;1/40;N;;;;;
0D5A;MALAYALAM FRACTION THREE EIGHTIETHS;No;0;L;;;;3/80;N;;;;;
0D5B;MALAYALAM FRACTION ONE TWENTIETH;No;0;L;;;;1/20;N;;;;;
0D5C;MALAYALAM FRACTION ONE TENTH;No;0;L;;;;1/10;N;;;;;
0D5D;MALAYALAM FRACTION THREE TWENTIETHS;No;0;L;;;;3/20;N;;;;;
0D5E;MALAYALAM FRACTION ONE FIFTH;No;0;L;;;;1/5;N;;;;;
0D76;MALAYALAM FRACTION ONE SIXTEENTH;No;0;L;;;;1/16;N;;;;;
0D77;MALAYALAM FRACTION ONE EIGHTH;No;0;L;;;;1/8;N;;;;;
0D78;MALAYALAM FRACTION THREE SIXTEENTHS;No;0;L;;;;3/16;N;;;;;
```

§7.2. Changes to NamesList.txt

```
@@ -59,4 +59,5 @@
 0D2D   MALAYALAM LETTER BHA
 0D2E   MALAYALAM LETTER MA
+      * also used to denote the fraction one eightieth (kaani)
 0D2F   MALAYALAM LETTER YA
 0D30   MALAYALAM LETTER RA

@@ -113,4 +114,22 @@
        * used alone to write the /au/ dependent vowel in modern texts
        x (malayalam vowel sign au - 0D4C)
+@              Minor fractions
+@+     Some minor fractions are represented by letters.
+*      The fraction one three-hundred and twentieth "muntiri" is denoted by the syllable
"pta" (0d2a 0d4d 0d24)
+*      The fraction one eightieth "kaani" is denoted by the letter "ma" (0d2e)
+0D58   MALAYALAM FRACTION ONE ONE-HUNDRED-AND-SIXTIETH
+      = arakaani
+0D59   MALAYALAM FRACTION ONE FORTIETH
+      = aramaa
+0D5A   MALAYALAM FRACTION THREE EIGHTIETHS
+      = muunnukaani
+0D5B   MALAYALAM FRACTION ONE TWENTIETH
+      = orumaa
+0D5C   MALAYALAM FRACTION ONE TENTH
+      = rantumaa
+0D5D   MALAYALAM FRACTION THREE TWENTIETHS
+      = muunnumaa
+0D5E   MALAYALAM FRACTION ONE FIFTH
+      = naalumaa
 @              Additional vowels for Sanskrit
 0D60   MALAYALAM LETTER VOCALIC RR

@@ -142,6 +161,15 @@
 @              Fractions
 0D73   MALAYALAM FRACTION ONE QUARTER
+      = kaal
 0D74   MALAYALAM FRACTION ONE HALF
+      = ara
 0D75   MALAYALAM FRACTION THREE QUARTERS
+      = mukkaal
+0D76   MALAYALAM FRACTION ONE SIXTEENTH
+      = maakaani
+0D77   MALAYALAM FRACTION ONE EIGHTH
+      = arakkaal
+0D78   MALAYALAM FRACTION THREE SIXTEENTHS
+      = muntaani
 @              Date mark
 0D79   MALAYALAM DATE MARK
```

§8. Code Chart

	0D0	0D1	0D2	0D3	0D4	0D5	0D6	0D7
0		ഐ	ഠ	ര	◌ീ		ൠ	൰
1			ഡ	റ	◌ു		ൡ	൱
2	◌ം	ഒ	ഢ	ല	◌ൂ		◌ൢ	൲
3	◌ഃ	ഓ	ണ	ള	◌ൃ		◌ൣ	൳
4		ഔ	ത	ഴ	◌ൄ			൴
5	അ	ക	ഥ	വ				൵
6	ആ	ഖ	ദ	ശ	െ◌		൦	൶
7	ഇ	ഗ	ധ	ഷ	േ◌	◌ൗ	൧	൷
8	ഈ	ഘ	ന	സ	ൈ◌	൘	൨	൸
9	ഉ	ങ	ഩ	ഹ		൙	൩	൹
A	ഊ	ച	പ		◌ൊ	൚	൪	ൺ
B	ഋ	ഛ	ഫ		◌ോ	൛	൫	ൻ
C	ഌ	ജ	ബ		◌ൌ	൜	൬	ർ
D		ഝ	ഭ	ഽ	◌്	൝	൭	ൽ
E	എ	ഞ	മ	◌ാ	ൎ	൞	൮	ൾ
F	ഏ	ട	യ	◌ി		ൟ	൯	ൿ

§9. Official Proposal Summary Form

(Based on N3902-F)

A. Administrative

1. Title

Proposal to encode Malayalam minor fractions

2. Requester's name

Shriramana Sharma

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

Individual Contribution

4. Submission date

2013-Apr-25

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

Malayalam

2. Number of characters in proposal

10 (ten)

3. Proposed category

Category B1, specialized small

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?

Shriramana Sharma

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

Shriramana Sharma. jamadagni-at-gmail-dot-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.

Yes.

C. Technical – Justification

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

This is a revision to my earlier document L2/13-051 with some small but important changes.

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?

Cibu Johny, Radhakrishna Warrier, Rajeesh Nambiar, Mahesh Pai

2c. If YES, available relevant documents

The matter was largely discussed in person or 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 who desire to store as digital text old Malayalam text involving these characters.

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?

Not at large. Perhaps by scholars.

5b. If YES, where?

Largely in research institutions around the world involved with Malayalam texts.

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?

The characters belong to the Malayalam block which is in the BMP and has sufficient space.

6c. If YES, reference

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

Yes, as far as possible. See §6 for details.

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?

Some characters are similar (but not identical) to existing Malayalam characters.

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

Yes

10c. If YES, reference

There is some similarity but the proposed characters have a consistently distinct shape.

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:

-o-o-o-