Request to encode Tamil fractions

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

Original: 2010-Sep-09, Revised: 2010-Oct-04

L2/10-334R

This document replaces L2/09-376, which requested the encoding of the three major fractions used in Tamil and Grantha in the Tamil block. L2/09-398 from INFITT and L2/09-416 from ICTA Sri Lanka had recommended that the major and minor Tamil fractions be together encoded in the SMP. In accordance with that, this document requests such an encoding in the SMP of fifteen fractions for Tamil, which includes both the major fractions one quarter, one half and three quarters, as well as minor fractions below one quarter.

§1. Tamil numerals and fractions to be re-used for Grantha

The Tamil digits 0-9 and Tamil numbers 10, 100 and 1000 are already encoded in the Tamil block in the BMP in the range 0BE6-0BF2. Despite their presence in the Tamil block and script=tamil property, they are used in Grantha texts also, as demonstrated in L2/09-372 p 5. Therefore, these same digits and numbers with the script=tamil property will be used along with Grantha characters (with script=grantha) being proposed separately, as recommended by L2/10-053 p 12 (though not in so many words). As there are already (at least) two precedents in:

1.	the Devanagari digits 0966-096F with script=devanagari being used with Kaithi (see also N3438), and

2.	the Bengali digits 09E6-09EF with script=bengali being used with Syloti Nagri

there should not be a problem in the 'Tamil' digits, which are in fact the Tamil-Grantha digits, being used with Grantha text as well. Therefore Grantha text should also use the Tamil numerals mentioned above which are already encoded, and the Tamil fractions which are being requested to be encoded in this document.

§2. List of fractions

The list of attested fractions is given below in ascending order with the Tamil names:

1.	௲	1/320	1/320	one three hundred and twentieth	muntiri / muntirai
2.	௹	2/320	1/160	one one hundred and sixtieth	araikkāṇi
3.	௺	4/320	1/80	one eightieth	kāṇi
4.	௱	8/320	1/40	one fortieth	araimā

5.	சூ	12/320	3/80	three eightieths	mukkāṇi
6.	ப	16/320	1/20	one twentieth	mā / orumā
7.	பஃ	20/320	1/16	one sixteenth	vīcam / mākāṇi
8.	உ	32/320	1/10	one tenth	irumā
9.	ஹௗ	40/320	1/8	one eighth	araikkāl
10.	௦	48/320	3/20	three twentieths	mummā
11.	௧ஃ	60/320	3/16	three sixteenths	mūvīcam / mummākāṇi
12.	௪	64/320	1/5	one fifth	nālumā
13.	வ	80/320	1/4	one quarter	kāl
14.	௨	160/320	1/2	one half	arai
15.	௬	240/320	3/4	three quarters	mukkāl

§3. Usage attestation

In L2/09-376 I had given attestation from ref 1 for the use of the major fractions. Attestation for the complete set of the major and minor fractions is provided by refs 2 and 3 which also clearly provide the numeric value of these fractions. (It is however noted that ref 2 omits 3/16 mummākāṇi.) These attestations are reproduced on pp 4-6 of this proposal.

§3.1. Fractions other (lesser) than those mentioned above

It is to be noted, however, that ref 3 gives written forms for fractional values even below 1/320, to be precise from 3/1280 (three one thousand two hundred and eightieths) called 'kīḻ mukkāl' down to 1/102400 (one one hundred and two thousand four hundredth) called 'kīḻ muntiri'. These are evidently obtained by multiplying the previous list of fractions by 1/320, the least among them. This is probably what caused L2/09-398 from INFITT and L2/09-416 from ICTA Sri Lanka to state* that there are "about 30" and "around 20" such fractions respectively.

However, there is no requirement to encode these fractions below 1/320 separately, since they are, as attested in ref 3, merely represented by the self-same fraction symbols mentioned above with a preceding 'kīḻ' or 'kī' (an abbreviation for the previous). There

---

* Further inquiries with INFITT elicited a response back in 2010-Jul-27 saying merely "INFITT will take care of Tamil fractions for Unicode" which has however not yet (as of 2010-Sep-08) taken fruit. The present author has also contacted ICTA Sri Lanka by email on 2010-Aug-31 regarding this and is yet (as of 2010-Sep-08) awaiting feedback.

should also be no requirement to atomically encode either 'kīḻ' or 'kī' as a multiplier equivalent to 1/320, because if the value meant by the sequence of the multiplier character followed by the fraction character can be understood by lexical analysis, then the same can be done with the consonant/vowel sequence 'kīḻ' or 'kī' followed by the fraction character.

I also note in passing that ref 3 mentions two even smaller fractions than 'kīḻ muntiri', viz 'immi' = 1/1075200 and 'aticāram' = 1/1838400 which do not appear to bear any instantly recognizable relationship with the 'base fraction' 1/320. However, distinct written forms for these two fractions are neither provided by ref 3 nor were (so far) obtainable from other sources. Therefore no separate characters for them are proposed.

The samples for attestation are placed on the following pages 4 to 6. For efficient use of space, I go on to the next section here.

§4. Unicode character properties

```
1xxx0;TAMIL FRACTION ONE THREE HUNDRED AND TWENTIETH;No;0;L;;;;1/320;N;;;;;
1xxx1;TAMIL FRACTION ONE ONE HUNDRED AND SIXTIETH;No;0;L;;;;1/160;N;;;;;
1xxx2;TAMIL FRACTION ONE EIGHTIETH;No;0;L;;;;1/80;N;;;;;
1xxx3;TAMIL FRACTION ONE FORTIETH;No;0;L;;;;1/40;N;;;;;
1xxx4;TAMIL FRACTION THREE EIGHTIETHS;No;0;L;;;;3/80;N;;;;;
1xxx5;TAMIL FRACTION ONE TWENTIETH;No;0;L;;;;1/20;N;;;;;
1xxx6;TAMIL FRACTION ONE SIXTEENTH;No;0;L;1xxx5 1xxx2;;;1/16;N;;;;;
1xxx7;TAMIL FRACTION ONE TENTH;No;0;L;;;;1/10;N;;;;;
1xxx8;TAMIL FRACTION ONE EIGHTH;No;0;L;;;;1/8;N;;;;;
1xxx9;TAMIL FRACTION THREE TWENTIETHS;No;0;L;;;;3/20;N;;;;;
1xxxA;TAMIL FRACTION THREE SIXTEENTHS;No;0;L;;;;3/16;N;;;;;
1xxxB;TAMIL FRACTION ONE FIFTH;No;0;L;;;;1/5;N;;;;;
1xxxC;TAMIL FRACTION ONE QUARTER;No;0;L;;;;1/4;N;;;;;
1xxxD;TAMIL FRACTION ONE HALF;No;0;L;;;;1/2;N;;;;;
1xxxE;TAMIL FRACTION THREE QUARTERS;No;0;L;;;;3/4;N;;;;;
```

The code points are obviously not allotted, pending the allocation of a Tamil Extended block as per the request of L2/09-317, and in consideration of the separately requested encoding of the characters for Extended Tamil from L2/10-256R.

Further, it should be noted here that the character for TAMIL FRACTION ONE SIXTEENTH has been given a canonical decomposition to that for TAMIL FRACTION ONE TWENTIETH followed by TAMIL FRACTION ONE EIGHTIETH because the symbol for one sixteenth is merely that for one twentieth followed by that for one eightieth. The separate encoding of this character is however justified by the fact that this fraction has a distinct name in Tamil – mākāṇi or vīcam, and that Tamil reference texts such as refs 2 and 3 hence list this fraction separately.

It is also recommended that the above order of characters in ascending order of absolute value of the fractions be followed since ascending order has been followed for the for the major fractions encoded elsewhere (i.e. 00BC..., 0B72..., 0D73..., A830... etc.)

From ref 1 p 273:
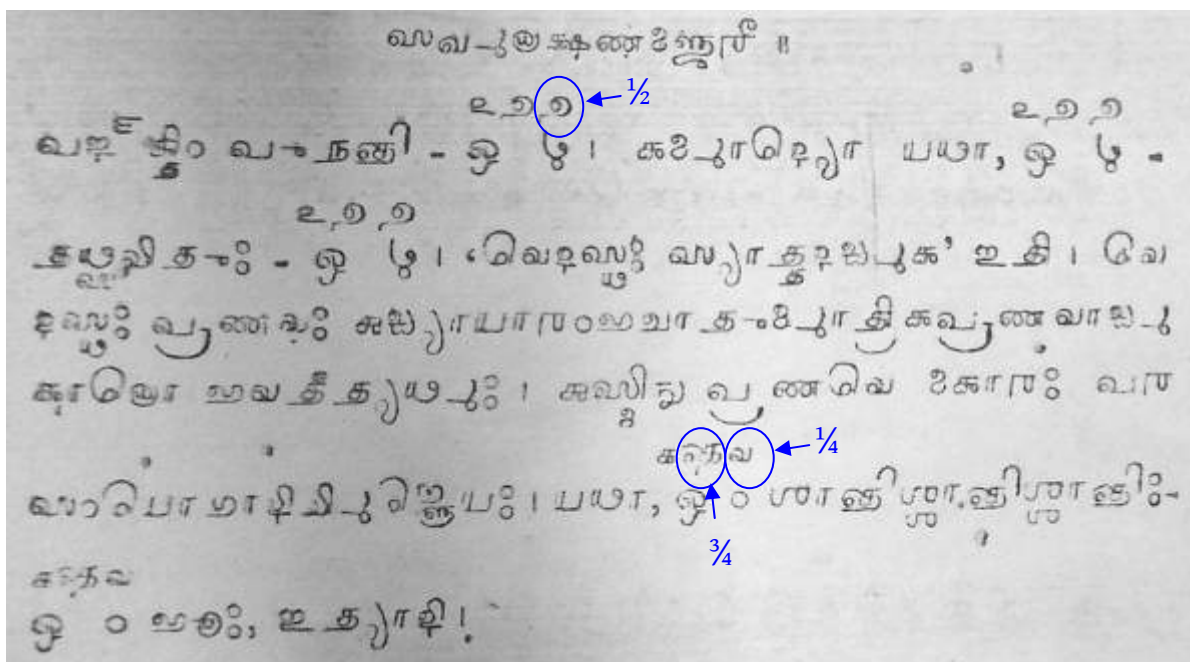
From ref 2 p 123:

Sign.	Name.	Power.	
ஸு	முந்திரை is the three hundred and twentieth part of the Integer,	$\frac{1}{320}$	
௳	அரைக்காணி	$\frac{1}{160}$	$\frac{2}{320}$
௴	காணி	$\frac{1}{80}$	$\frac{4}{320}$
௅	அரைமா	$\frac{1}{40}$	$\frac{8}{320}$
௴	முக்காணி	$\frac{3}{80}$	$\frac{12}{320}$
�	மா or ஒருமா	$\frac{1}{20}$	$\frac{16}{320}$
௴	மாகாணி or வீசம்	$\frac{1}{16}$	$\frac{20}{320}$
௨	இரண்மா or இருமா	$\frac{1}{10}$	$\frac{32}{320}$
௲	அரைக்கால் or இரண்மாவரை	$\frac{1}{8}$	$\frac{40}{320}$
௳	மும்மா or மூன்றுமா	$\frac{3}{20}$	$\frac{48}{320}$
௳	நாலுமா	$\frac{1}{5}$	$\frac{64}{320}$
�வ	கால்	$\frac{1}{4}$	$\frac{80}{320}$
௨	அரை	$\frac{1}{2}$	$\frac{160}{320}$
௴	முக்கால்	$\frac{3}{4}$	$\frac{240}{320}$

	Fraction	Tamil
க	1	ஒன்று
ஃ	3/4	முக்கால்
?	1/2	அரை
?	1/4	கால்
?	1/5	காலுமா
?	3/16	அரைக்கால்
உ	1/10	இருமா
ரூ	3/16	மும்மாகாணி
எ	1/16	மாகாணி, வீசம்
அ	3/80	மும்மா
?	1/20	ஒருமா
?	1/40	அரைமா
?	3/80	முக்காணி
?	1/80	காணி
?	1/160	அரைக்காணி

		Fraction	Tamil
	?	1/320	முந்திரி
கீழ்	ஃ	9/1280	கீழ்முக்கால்
கீ	?	1/640	கீழ்மா
கீ	வ	1/1280	கீழ் கால்
கீ	?	1/1600	கீழ் காலுமா
கீ	ரூ	3/5120	கீழ் மூன்றுவீசம்

ஜ	௸	$\dfrac{3}{6,400}$	இழ் மும்மா
ஜ	ௗ	$\dfrac{1}{2,800}$	இழ் அரைக்கால்
ஜ	உ	$\dfrac{1}{3,200}$	இழ் இருமா
ஜ	பத	$\dfrac{1}{5,120}$	இழ் வீசம்
ஜ	ப	$\dfrac{1}{6,400}$	இழொருமா
ஜ	சூ	$\dfrac{3}{25,600}$	இழ் முக்காணி
ஜ	சு	$\dfrac{1}{12,800}$	இழரைமா
ஜ	ஐ	$\dfrac{1}{25,600}$	இழ்க்காணி
ஜ	௰	$\dfrac{1}{6,12,000}$	இழ் அரைக்காணி
ஜ	௯	$\dfrac{1}{1,02,400}$	இழ் முக்கிரி
இம்மி		$\dfrac{1}{10,75,200}$	
அதிசாரம்		$\dfrac{1}{18,36,400}$	

§5. References

1. Vyāsa Śikṣā with Veda Taijasa commentary, Brahmananda Press, Tiruvadi, 1908.

2. A Grammar of the Common Dialect of the Tamul Language called Kotuntamiḻ, Constantius Joseph Beschi, Tr: George William Mahon, Madras, 1848.

3. Iniya Tamiḻ Ilakkaṇam, Yoki Śrī Cuttānanta Pāratiyār, Kavita Publications, T Nagar, Chennai; http://tamilelibrary.org/teli/numeral.html, retrieved 2010-Sep-08

Note on revised version of document:

The previous version of this document L2/10-334 had listed some of the fractions in the wrong order, despite specifying that the encoding should be done in the listed order in ascending order of absolute value. I apologise for this serious error and submit this revised version which ensures that the fractions are now in the correct order by showing (in §2) their numerators to the 'base' denominator 320. This revised version also includes a code chart and accompanying description which should also ideally be part of a proposal.

§6. Code chart

Description:

	1xxx
0	ஒஃகூ
1	௰
2	௱
3	சு
4	கூ
5	ப
6	பஃ
7	உ
8	ஹ௫
9	௸
A	௹
B	சு
C	வ
D	௲
E	௫
F	

1xxx0 ஒஃகூ TAMIL FRACTION ONE THREE HUNDRED AND TWENTIETH
= muntiri

1xxx1 ௰ TAMIL FRACTION ONE ONE HUNDRED AND SIXTIETH
= araikkaani

1xxx2 ௱ TAMIL FRACTION ONE EIGHTIETH
= kaani

1xxx3 சு TAMIL FRACTION ONE FORTIETH
= araimaa

1xxx4 கூ TAMIL FRACTION THREE EIGHTIETHS
= mukkaani

1xxx5 ப TAMIL FRACTION ONE TWENTIETH
= maa

1xxx6 பஃ TAMIL FRACTION ONE SIXTEENTH
= viicam or maakaani

1xxx7 உ TAMIL FRACTION ONE TENTH
= irumaa

1xxx8 ஹ௫ TAMIL FRACTION ONE EIGHTH
= araikkaal

1xxx9 ௸ TAMIL FRACTION THREE TWENTIETHS
= mummaa

1xxxA ௹ TAMIL FRACTION THREE SIXTEENTHS
= muuviicam or mummaakaani

1xxxB சு TAMIL FRACTION ONE FIFTH
= naalumaa

1xxxC வ TAMIL FRACTION ONE QUARTER
= kaal

1xxxD ௲ TAMIL FRACTION ONE HALF
= arai

1xxxE ௫ TAMIL FRACTION THREE QUARTERS
= mukkaal

§7. Official Proposal Summary Form

A. Administrative

1. Title
Request to encode Tamil fractions
2. Requester's name
Shriramana Sharma
3. Requester type (Member body/Liaison/Individual contribution)
Individual contribution
4. Submission date
Original: 2010-Sep-09, **Revised**: 2010-Oct-04
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, except for the actual code points which should be allotted by the UTC.

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. **This is a proposal for fractions that are used with the Tamil and Grantha scripts**.
1b. The proposal is for addition of character(s) to an existing block, Name of the existing block
Block not yet existing but requested by L2/09-317. Name of block suggested to be Tamil Extended.
2. Number of characters in proposal
15 (fifteen)
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?
Elmar Kniprath.
b. Identify the party granting a license for use of the font by the editors (include address, e-mail etc.)
Elmar Kniprath. kniprath – at – online – dot – de.
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes, for all the fractions proposed.
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes, for the major fractions.
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.
A proposal for addition of the Tamil major fractions alone was submitted as L2/09-376 but later it was decided to encode all the Tamil fractions (major and minor) based on feedback L2/09-398 from INFITT, L2/09-416 from ICTA Sri Lanka etc.
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. The proposer himself is also a member of the user community.

2b. If YES, with whom?

National bodies: Ministry of IT, Govt of India. ICTA, Sri Lanka. Experts: Dr Reinhold Grünendahl, Germany. Dr T Ganesan, Institut Français de Pondichéri. Dr Jean-Luc Chevillard, France. Users: Vinodh Rajan, Sriramadas Mahalingam and Ramanraj K, all from Chennai. Other: INFITT.

2c. If YES, available relevant documents

None specifically. The matter was discussed personally/via email.

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

Tamil is spoken and written by a large user community in Tamil Nadu and elsewhere. The Grantha script is used for Sanskrit by Vedic scholars and others reading/writing Sanskrit in Tamil Nadu.

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

Rare usage in contemporary Tamil books. Sometimes seen in Sanskrit works in Grantha.

4b. Reference

See proposal.

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

Yes. Especially the major fractions. Usage somewhat rare, however.

5b. If YES, where?

Tamil Nadu, India.

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No.

6b. If YES, is a rationale provided?

6c. If YES, reference

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

Yes, since it is only logical to keep mutually related characters together.

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?

Yes, TAMIL FRACTION ONE SIXTEENTH = TAMIL FRACTION ONE TWENTIETH + TAMIL FRACTION ONE EIGHTIETH.

9b. If YES, is a rationale for its inclusion provided?

Yes.

9c. If YES, reference

The rationale is that this fraction has a distinct name in Tamil and is hence listed separately in Tamil reference texts. Its equivalence to the sequence can however be handled by canonical decomposition.

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Yes, some characters are identical with and others similar to Tamil or Grantha consonant letters.

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

Yes.

10c. If YES, reference

Their Unicode character properties would differ.

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?

No.

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)?

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

If YES, reference:

No.

–o–o–o–