

Revised Tamil fractions proposal

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

2010-Oct-22

This document requests two deviations from the Tamil fractions proposal L2/10-334R based on fresh information obtained after the submission of that document. As such it effectively supersedes that document.

In L2/10-334R it was requested that TAMIL FRACTION ONE SIXTEENTH be encoded separately though it is merely a sequence of the fractions ONE TWENTIETH and ONE EIGHTIETH:

$$\text{பு} (1/16) = \text{பு} (1/20) + \text{ஐ} (1/80)$$

The justification given was that this fraction has a distinct name in the Tamil language viz *vīcam* and is hence listed separately in Tamil reference texts. To preserve the glyphic and semantic equivalence it was suggested that there be a canonical decomposition of ONE SIXTEENTH to the sequence of ONE TWENTIETH and ONE EIGHTIETH.

However, some sources suggest that TAMIL FRACTION ONE SIXTEENTH has also been written in a ligature-like form. We provide a sample from p 58 of the reference work *South Indian Scripts in Sanskrit Manuscripts and Prints* by Reinhold Grünendahl (ISBN 3-447-04504-3), published in 2001 in Wiesbaden, Germany:

வெ

$$1/320$$

நி

$$1/160 = 2/320$$

ஐ

$$1/80 = 4/320$$

சு & சு

$$1/40 = 8/320$$

சாநி

$$1/32 = 8/320 + 2/320$$

சு

$$3/80 = 8/320 + 4/320$$

பு

$$1/20 = 16/320$$

பு & வெ

$$1/16 = 16/320 + 4/320$$

ஐ

$$1/10 = 32/320$$

ஐ & ஐ

$$1/8 = 40/320$$

நி

$$3/20 = 48/320$$

சு

$$1/5 = 64/320$$

வெ

$$1/4 = 80/320$$

ஐ & வெ

$$1/2 = 160/320$$

ஐ

$$3/4 = 240/320$$

However, the other references that list the Tamil fractions do not attest to this ligature-like form. See L2/10-334R pp 4-5. Further inquiry from the author of the above reference also did not produce conclusive evidence that this ligature-like form has indeed been consistently used in manuscripts or prints to be worthy of immediate encoding. If any evidence for the consistent use of this ligature-like form for one sixteenth is found in future, only then would it be justified to encode it as TAMIL FRACTION ONE SIXTEENTH.

As for ONE SIXTEENTH being written as the sequence of the glyphs of ONE TWENTIETH and ONE EIGHTIETH, it is always possible to represent this written form in encoding by a sequence of the two separate characters of ONE TWENTIETH and ONE EIGHTIETH. So:

$$\text{TAMIL FRACTION ONE TWENTIETH} + \text{TAMIL FRACTION ONE EIGHTIETH} = \text{ூ௩} = 1/16$$

This is in accordance with the additive nature of the old-style Tamil numbers. Thus there is no requirement to encode a separate character for the glyphic sequence form especially seeing as it would then be required to canonically decompose it to the character sequence. If in the future sufficient indubitable attestation is found for the ligature-like form of one sixteenth noted above, then it would only be appropriate that that ligature-like form be encoded as the distinct character TAMIL FRACTION ONE SIXTEENTH.

The fact that such an encoding of the ligature-like form in the future would make it possible to represent the fractional value one sixteenth in two different ways in encoding is also not to be found fault with. The fraction three eightieths (mukkāṇi) can already (that is, as currently otherwise proposed) be represented in two ways: as a sequence of the characters ONE FORTIETH and ONE EIGHTIETH and as the distinct ligature-like form represented by the character THREE EIGHTIETHS:

$$\text{TAMIL FRACTION THREE EIGHTIETHS} = \text{௪௩} = 3/80$$

$$\text{TAMIL FRACTION ONE FORTIETH} + \text{TAMIL FRACTION ONE EIGHTIETH} = \text{௪௩} = 3/80$$

Comparing the two written forms above it appears that the first form (which we have termed ligature-like) may indeed have developed as a ligated version of the latter. However, it is only advisable to encode the ligature-like form as the separate character THREE EIGHTIETHS since it is inadvisable to bring in ZWJ (needed for requesting ligatures) into numbers. So the fact remains that three eightieths has two encoded representations with the existing proposal. Thus if in the future sufficient attestation is found for the ligature-like form of one sixteenth it would entirely be justified to encode it.

For the present, this means that no character ONE SIXTEENTH should be encoded. The place should however be kept reserved with an annotation to use the sequence of the characters ONE TWENTIETH and ONE EIGHTIETH.

Further, it is noted that the reference above lists one by thirty-two as a distinct fraction. However the written form provided is just the sequence of the glyphs of ONE FORTIETH and ONE ONE HUNDRED AND SIXTIETH. This can be represented in encoding by the sequence of the corresponding characters. It is however possible that further research may turn up a ligature-like form just as has been mentioned for one sixteenth. This fractional value one by thirty-two also has a name in Tamil – araiṁcam (lit.: half of ṁcam). Thus it may be necessary in the future to encode a character TAMIL FRACTION ONE THIRTY SECOND.

In view of this, it is recommended that in the series of fraction characters, there be reserved a space between the character ONE FORTIETH ($1/40 = 8/320 = 0.025$) and THREE EIGHTIETHS ($3/80 = 12/320 = 0.0375$) for a possible future character ONE THIRTY SECOND ($1/32 = 10/320 = 0.03125$) to preserve the ascending order of absolute value.

We note that this argument might be further extended to recommend a space between ONE EIGHTIETH and ONE FORTIETH for a possible character ONE SIXTY FOURTH (which would be called kālṁcam in Tamil). However, as no references obtained by us so far list such a fraction distinctly even as a glyphic sequence, we refrain from doing so.

Revised 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;reserved
1xxx5;TAMIL FRACTION THREE EIGHTIETHS;No;0;L;;;;3/80;N;;;;;
1xxx6;TAMIL FRACTION ONE TWENTIETH;No;0;L;;;;1/20;N;;;;;
1xxx7;reserved
1xxx8;TAMIL FRACTION ONE TENTH;No;0;L;;;;1/10;N;;;;;
1xxx9;TAMIL FRACTION ONE EIGHTH;No;0;L;;;;1/8;N;;;;;
1xxxA;TAMIL FRACTION THREE TWENTIETHS;No;0;L;;;;3/20;N;;;;;
1xxxB;TAMIL FRACTION THREE SIXTEENTHS;No;0;L;;;;3/16;N;;;;;
1xxxC;TAMIL FRACTION ONE FIFTH;No;0;L;;;;1/5;N;;;;;
1xxxD;TAMIL FRACTION ONE QUARTER;No;0;L;;;;1/4;N;;;;;
1xxxE;TAMIL FRACTION ONE HALF;No;0;L;;;;1/2;N;;;;;
1xxxF;TAMIL FRACTION THREE QUARTERS;No;0;L;;;;3/4;N;;;;;
```

Revised code chart

1xxx	1xxx0	ஹு	TAMIL FRACTION ONE THREE HUNDRED AND TWENTIETH = muntiri
0 ஹு	1xxx1	ற	TAMIL FRACTION ONE ONE HUNDRED AND SIXTIETH = araikkaani
1 ற	1xxx2	ஐ	TAMIL FRACTION ONE EIGHTIETH = kaani
2 ஐ	1xxx3	சு	TAMIL FRACTION ONE FORTIETH = araimaa
3 சு	1xxx4		<reserved for one thirty second> = araiviicam
4	1xxx5	சூ	TAMIL FRACTION THREE EIGHTIETHS = mukkaani
5 சூ	1xxx6	ப	TAMIL FRACTION ONE TWENTIETH = maa
6 ப	1xxx7		<reserved for one sixteenth> = viicam or maakaani * written as பஐ encoded as 1xxx6 1xxx2
7	1xxx8	ஊ	TAMIL FRACTION ONE TENTH = irumaa
8 ஊ	1xxx9	ஹு	TAMIL FRACTION ONE EIGHTH = araikkaal
9 ஹு	1xxxA	ம	TAMIL FRACTION THREE TWENTIETHS = mummaa
A ம	1xxxB	மூ	TAMIL FRACTION THREE SIXTEENTHS = muuviicam or mummaakaani
B மூ	1xxxC	சு	TAMIL FRACTION ONE FIFTH = naalumaa
C சு	1xxxD	வ	TAMIL FRACTION ONE QUARTER = kaal
D வ	1xxxE	உ	TAMIL FRACTION ONE HALF = arai
E உ	1xxxF	ஐ	TAMIL FRACTION THREE QUARTERS = mukkaal
F ஐ			

Revised Official Proposal Summary Form

A. Administrative

1. Title

Revised Tamil fractions proposal

2. Requester's name

Shriramana Sharma

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

Individual contribution

4. Submission date

2010-Oct-22

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

14 (fourteen)

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. L2/10-334R was then submitted. Now a small change is requested.

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, Chennai; Elmar Kniprath, Germany. 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?

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?

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)?

No.

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

If YES, reference:

-O-O-O-