

Reply to L2/14-048 on Tamil fractions

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

2014-Jul-18

This is with regard to L2/14-048 “Comments on the Proposals to Encode Tamil Symbols and Fractions” submitted by the ICTA Sri Lanka.

Handling fractions using font substitution

ICTA’s document suggests that Tamil fractions should not be encoded as atomic characters but should be handled by substitution rules in fonts as follows:

ஃ/௩௨ → ஃ௩௨

i.e.

OBE7 TAMIL DIGIT ONE + 2044 FRACTION SLASH* + OBE9 TAMIL DIGIT THREE + OBE8 TAMIL DIGIT TWO
→ TAMIL FRACTION ONE THIRTY-SECOND

However Unicode encodes characters based on their written form, not based on their meaning. Therefore written forms which are *not* presentation variants of or otherwise orthographically derived from the nominal glyphs of one or more characters are *not* handled by font substitution rules in Unicode.

Fractions in international format like $1/32$ are indeed just a presentation form of the components 1, 3 and 2. However this is not the case in Tamil fractions. For instance, ஃ௩௨ (with value $1/32$) is not a presentation variant of or orthographically derived from the shapes of the nominal components ஃ (1) and ௩௨ (32).

As such, it is not appropriate to represent the uniquely shaped Tamil fractions such as ஃ௩௨ using a sequence of the regular Tamil digits by font substitution.

The sequence suggested above would in fact be used for something like: $\overset{\text{ஃ}}{\text{௩௨}}$ (which uses the Tamil digits for 1, 3 and 2 in the modern format for fractions).

* The document doesn’t specifically mention U+2044 FRACTION SLASH against U+002F SOLIDUS but it seems it would be the logical choice for such a sequence.

Concerns about “never-ending additions”

ICTA voices concern that encoding individual fractions may “lead to a never-ending set of additions” since “some other fractions than the proposed fractions may exist”.

However, it is a goal of Unicode to provide a standardized encoding to help textually digitize such ancient texts also and not only texts generated by contemporary usage. Thus one does not shie away from encoding an attested character just because it is possible that there are other such attested characters. Given this, if historical documents turn up further fractions which are not merely glyphically composed of other numerals, then certainly they should be considered for separate encoding.

I should also point out that back in 2009, it was the ICTA which spoke out against my proposal L2/09-376 to encoding the Tamil major fractions *alone* in the BMP, and wrote as follows in L2/09-416 (underlining mine):

- 1) *There are many more numerals (fractions) which should be encoded in Tamil Unicode. We have currently identified around 20 such fractions, and believe there may be a few more.*
- 2) *We recommend that all the known Tamil fractions be encoded together (with spaces left for any others which may be identified later), rather than just the 3 major fractions as requested in L2/09-376.*

There is no need for the ICTA to take a new stance now especially since there are no technical faults with its older stance which aligns quite well with the standing proposal.

Thus there need be no hesitation to encode these fractions which are already approved by the UTC and WG2 committees at their meetings in 2013 May and June.