

Feedback on the GoTN's document L2/16-062 on Tamil fractions and symbols

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

2016-May-02

With regard to my original proposal L2/12-231 and revised proposal L2/13-047 N4430, the Gov't of Tamil Nadu (henceforth, "GoTN") has submitted its feedback as L2/16-039 and (superseding that) as L2/16-062. The following are my comments on L2/16-062.

§1. Characters retracted earlier but are now re-proposed

In my revised proposal L2/13-047 N4430 p 1, I listed the characters pairs from my original proposal which I had sought to disunify but which the review committee and UTC recommended to *not* disunify. The excess characters were thus removed from the revised proposal. Now the GoTN presents evidence for some of these characters as being glyphically distinct from those they are unified with and asks to separately encode them.

§1.1. Fraction One Twentieth vs Letter PA

The printed evidence from my original proposal showed the same glyph \sqcup as consistently being used for both the letter PA and the fraction one-twentieth. On p 17, the GoTN doc requests that a rounded variant be encoded as a separate character for the fraction.

On this p 17, the GoTN doc successfully shows contrastive attestation between an angular form for the letter PA and a rounded shape for the fraction. No doubt, the later printings I had examined in my proposal had conflated the two. As such, it would seem that there is indeed justification for encoding this as a separate character.

Note that on the previous page (p 15) the GoTN doc claims that the fraction also has a taller right arm: \cup . However, the attestation provided only shows arms of equal length:



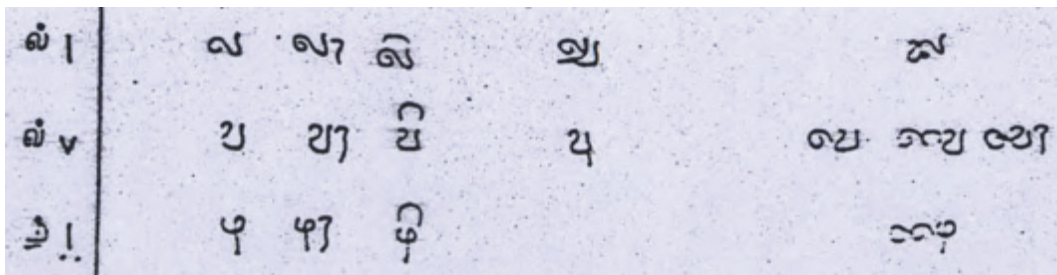
Finally, the evidence provided here is not really that of an independent usage of this sign \cup for this fraction but is the same evidence provided on p 11 for the ligature $\cup\text{C}$ denoting the fraction one-sixteenth where this sign is seen as the leading part. Presumably,

the intent is to provide grounds for *inferring* the presence of an independent (i.e. outside ligatures) usage of ூ as the fraction one-twentieth. Since the inference seems justified, I would support encoding such a character.

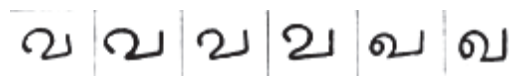
§1.2. Fraction One Quarter vs Letter VA

The printed evidence from my original proposal showed the same glyph ூ as consistently being used for both the letter VA and the fraction one-fourth. On p 16, the GoTN doc requests that a variant of the same with a higher “head” be encoded as a separate character for the fraction: ூ. However, there is no evidence provided (unlike that was given above for the rounded vs angular PA) for this to contrast with ூ for the letter VA.

In fact, it would seem that the letter VA in those days was itself written with a higher “head” than today (where it rests on the baseline). If I read the header line of the hazy scan correctly, the sample provided is from an inscription of Parthivendra Varman who is identified with Aditya Chola, the elder brother of Rajaraja Chola I (985-1014 CE). From p 50 of *Evolution of script in Tamil Nadu* (R Govindaraj, pub. Tamil Nadu Archaeological Society, 1994), I present the shape of VA from a 986 CE inscription of Uttama Chola:






This clearly shows a higher head than the ூ of today and the only remaining difference is the presence or absence of the loop in the head but this is not an issue, especially since the epigraphist source Caa Ganesan quoted on p 33 of my original proposal L2/12-231 clearly derives the current form ூ from this epigraphical form without the loop:









As such, the evidence still seems to be insufficient to disunify the characters.



§1.3. Sign Uzhakku vs Fraction Three Quarters





The printed evidence from my original proposal showed the same glyph ூ as consistently used for both the fraction three-quarters and the grain measure *uḷakku*. On p 31, the GoTN doc requests that a variant without the loop be encoded separately for the measure: ூ.


However, there is no direct evidence presented for the independent usage of this glyph. The only evidence presented is as part of the ligature  denoting 3 × *ulakku* where the leading part  is the Tamil digit 3. (Note that we have separately proposed this ligature with the glyph  and it stands as 11FD5 TAMIL SIGN MUUVULAKKU in the latest draft 5th ed CD2 L2/15-339 N4705).

Since it is normal for ligatures to elide strokes, even with a representative glyph *with* the loop, the ligature can elide the loop:  +  ->  or . As such, without direct evidence of consistent usage of  as the sign for *ulakku* in contrast with  as the fraction three-quarters, it would seem difficult to justify encoding it a separate character.


§1.4. Sign Kuruni vs Letter NGA

The printed evidence from my original proposal showed the same glyph  as consistently being used for both the letter NGA and the grain measure *kuruni*. On p 32, The GoTN doc requests that an entirely different glyph  be encoded as the character for *kuruni*.

Evidence for its usage is provided but no formal reference texts identifying this sign as that of *kuruni* are quoted. There is no resemblance between this and the sign  that is clearly identified by printed publications from native sources such as Kanakkadigaram 1958 and Cintamani En Suvadi (pp 39, 41 of L2/12-231) even if anyone were to suspect non-native authors such as Beschi and Winslow (ibid pp 40, 46) of somehow having been misled. Beschi even shows a *ligature* of the digit two  with  to produce .

As such, it would seem prudent to seek further correlation and consistent usage before encoding  as a character for *kuruni*.

§1.5. Recommendation

It seems justified to encode  as the fraction for one twentieth since evidence of usage, even if only as part of a ligature, is provided and in contrast with the angular shape used in the same sources for the letter PA. If the UTC agrees, I shall separately submit a formal proposal (since the GoTN doc lacks character properties).

For the other characters, the evidence provided is insufficient to justify separate encoding. As the UTC remarked back in 2013 Feb, these can always encoded later if sufficient evidence of consistent and contrastive usage is available.

§2. On the remaining characters

Apart from the above, the other characters have been deliberated and voted upon by the committees at various stages. So the first task at hand is to examine whether there is a problem with encoding them as they are on ballot. We shall proceed to do that now.

Note except for the broad categories of fractions, currency signs etc, the GoTN doc does not seem to follow any particular order in discussing individual characters *within* the categories. Now since we have first had to discuss characters which are *not* on the ballot in the previous section, the order (whatever it may be) from the GoTN doc is already broken. As such, for convenience of the committees in evaluating whether or not a change is justified to the characters already on ballot, we shall proceed in codepoint order. For the same reason, the codepoints below refer to the draft 5th ed CD2 L2/15-339 N4705.

Note that the GoTN doc also asks for sweeping changes in the naming policy, spelling and annotations of the characters. We shall first discuss only the glyphic identity which is the core of the characters. A later section will consider the other changes.

§3. Fractions

§3.1. 11FC2 ெ ONE EIGHTIETH





The glyph currently proposed is based on a variant found in Shuddhananda Bharati, Kanakkadigaram 1958 and Cintamani En Suvadi (pp 30, 35-38 of L2/12-231). The GoTN requests on pp 12-14 that a variant with a rounded finish to the bottom right ெ should be chosen. This is also found attested in some of the printed sources I had quoted such as Kanakkadigaram 1928, Beschi, Beythan, Winslow and Wickremasinghe. I had simply chosen ெ to be distinctive from 1132E ெ GRANTHA LETTER MA to minimize confusability issues. However, I realize that these characters are not going to be part of IDNs. As such, I have no problem with the representative glyph being modified to ெ.

§3.2. 11FC5 ெ ONE THIRTY-SECOND

On p 24, the GoTN doc says that the “derivation” found in §4.4 of L2/12-231 is “unacceptable” as it is “incorrect”. It presumes that I had chosen to “derive” it this way “as there was not any valid symbol could [sic] be found in the sources” and comments that the symbol has been “forcefully matched with an available collection”.

Firstly, the doc seems to have misunderstood the purport of §4.4 of L2/12-231. I have not *artificially* “derived” any glyph for this or any other characters. I have only posited *probable* derivations for *attested glyphs* which *prima facie* follow from the available evidence as explained in that section of my original proposal.

The evidence for this particular glyph is provided on L2/12-231 p 38 from a book published in 1950 by C Kumarasamy Naidu & Sons (34, Chinnathambi Street, Chennai). This is not evidence that I have concocted. I can arrange to provide the entire scanned book if there is any doubt about its authenticity. As such, I do not see what is there to “forcefully match” here. Why is this evidence only “an available collection” and not considered a valid source, so much so that it is presumed “there was not any valid symbol ... in the sources”?

On top of this, the GoTN doc provides a symbol  without any attestations at all. It claims that this is “derived” from half + one-sixteenth (since $\frac{1}{2} \times \frac{1}{16} = \frac{1}{32}$). On what basis is this “derivation” done is not explained. Further, while the glyph  for one-sixteenth is indeed similar to the trailing part of this glyph, the leading part of this glyph is unrelated to both the variants observed for one-half:  and . As such, it is not clear how attempting to strike out an attested glyph in favour of an unattested one is justified.

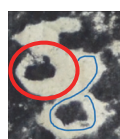
§3.3. 11FC6 சு Three Eightieths

Just as for 11FC2 ஶ ONE EIGHTIETH discussed above, the glyph currently proposed for three eightieths is based on a variant found in some sources whereas other sources show a shape with a rounded finish to the bottom right. In §4.4 of L2/12-231 I had observed that this character appears to be a ligature of the summative component fractions:

$$mukkāṇi \frac{3}{80} = araimā \frac{1}{40} + kāṇi \frac{1}{80} = சுஶ \rightarrow சு$$

Keeping in with ஶ for $\frac{1}{80}$, I had proposed சு for $\frac{3}{80}$. If ஶ is to be chosen for the former, it is quite appropriate to choose ஶ for the latter. However, on p 14 the GoTN says that the “correct symbol” is ஶ where in addition to the rounded finish, the leading part is that of the digit three ௩, thus losing the curved connection on the left: ஶ vs ஶ.

However, the sample from the estampage provided on p 13 actually shows a curved connection on the left (red circle below) which goes in favour of ஶ and not ஶ:

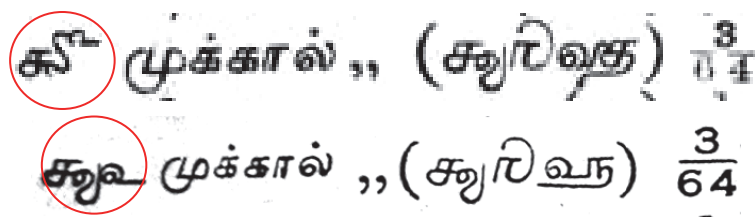


The idea to use a 𑌞-like shape as part of the ligature with 𑌚 seem to be based on the transcription of the inscription (above right). However, this is not direct evidence for the shape 𑌚, specifically of the 𑌞 part to its top left. The preference to use a rounded stroke in the bottom right can be accommodated however and the glyph may be changed to 𑌚𑌞.

The GoTN recommends to postpone the encoding of this character. However, sufficient evidence is available for the presence of the character and even for its currently proposed glyph as shown from multiple sources in the original proposal L2/12-231 p 29, 30, 34, 35, 37 and 38. If evidence is provided later for a *more consistently used* glyph (which I personally consider highly unlikely), then a glyph change can be effected. I thus do not see the need for postponing this character.

§3.4. 11FC7 𑌚𑌞 THREE SIXTY-FOURTHS

On p 25, the GoTN quotes the samples provided for this glyph on pp 35 and 38 of L2/12-231:



It further observes that “the symbols in both differ a lot”. I would have thought that it does not require too much a conceptual stretch for the unligated form seen in the lower sample to lead to the ligated form in the higher sample (red circles above). The GoTN doc also says that “multiple fractions occur in sequence, but combining all of them to make another symbol is incorrect”, that this looks (again) like “matching a symbol in the attestation forcefully” and that “the panel of GoTN doesn’t suggest this way of creating a symbol”.

Again, just as in the case of 11FC5 𑌚𑌞 ONE THIRTY-SECOND above, is the implication that this shape has been “*created*” by me? Given the attestation, such a statement would be baseless. Is then the the authenticity of the attestation provided called into question again? What is “forceful” about this identification? Even the alternate glyph 𑌚𑌞 identified by the GoTN allows for interpretation as a ligature based on the leading fraction in the sequence: 𑌚. Such a glyph variant can always be taken care of by a font appropriate to the particular inscription or period or region, since the rest of the Tamil characters such as letters and vowel signs will also certainly have their own unique shapes in those inscriptions.

The GoTN doc has not explained why exactly the provided evidence is unpalatable. As such, I do not see sufficient grounds to remove this character from the ballot. Note that later on in this document I agree with the GoTN on removing some characters from the ballot. As such, it should be clear that I have no stubborn insistence on encoding these just because I proposed them.

In fact, the presence of the form ஈ) identified by the GoTN’s committee clearly indicates that this character exists, whatever be its representative glyph. If in future, evidence should be provided that that glyph is *more* consistently used than the one chosen here: ஈ, then there will be no problem in changing the glyph to a “better” one.

§3.5. 11FC8 ஸ ONE SIXTEENTH - Current character

On p 9, the GoTN requests that the representative glyph of this character be changed to ூ where “the height of the ridge is not same as that of the arms” and which is “slightly tilted anticlockwise”. No direct attestation is provided for consistent use of such a very specific shape, but reference is made to the Caa Ganesan attestation in my original proposal L2/12-231 p 33. However, that attestation shows all kinds of historical variants:



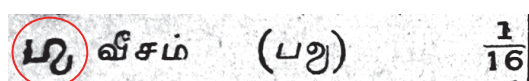
Further, if an epigraphic shape for this characters is to be adopted as the representative glyph “to reflect the natural usage”, it would imply that the glyphs for other characters should be changed to epigraphic shapes as well which would certainly be unacceptable. The proposed glyph is from a source (L2/12-231 p 35) that uses the modern Tamil script:



As such, stylistic consistency would dictate that the current glyph is not changed.

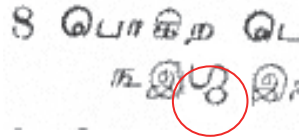
§3.6. ONE SIXTEENTH - Additional character

On p 11, the GoTN further requests that a different written form of the same fraction is encoded as a separate character: ூ. This is nothing but the rounded version of the alternate rendering of this fraction: ஸ which I had documented in L2/12-231 p 38:



Sufficient attestation is provided for the rounded shape (estampage and transcription):





I have noted in L2/12-231 pp 9, 10 §4.4.2 that this shape is evidently the ligature of $\sqcup m\bar{a}^{1/20}$ with $\text{ᯚ} k\bar{a}n\bar{i}^{1/80}$ to give $^{1/20} + ^{1/80} = ^{1/16}$. As noted previously in §1.1 and §3.1, the GoTN requests that a rounded form of \sqcup i.e. \cup be encoded separately for $m\bar{a}$ and that the glyph for $k\bar{a}n\bar{i}$ should be slightly changed to ᯚ . The ligature of \cup with ᯚ would certainly be ᯚ .

While both ᯚ and \cup denote the same mathematical value and are superficially similar in appearance, one can argue that their character identities are distinct as the former is a ligature whereas the latter seems to be an *independent* historical alternate representation as evidenced by the above sources.

Since I had insufficient evidence to be convinced about the “independent” part, I had not proposed to disunify this character in L2/12-231. Even now, the evidence seems insufficient to establish that \cup , though historically attested, is not just a simplification of the ligature ᯚ . Further evidence of contrastive usage of the two forms would be preferable but this seems very unlikely since they both represent the same value.

While the GoTN doc (on p 11) expresses concerns over “wrong recording/leading of old texts” and wishes to ensure that “encoding doesn’t deviate from history”, one should remember that plaintext and hence the encoding cannot be expected to represent the exact appearance of inscriptions. Since contrastive usage within the same context has not been established, it is difficult to assert that the two are different characters. The scholars the GoTN consulted seem to have felt differently, though. I leave it to the UTC to decide.

§3.7. 11FCA ᯚ ONE EIGHTH

On p 20, the GoTN requests that the glyph be changed to ᯚ to reflect a variant seen in one particular publication, the presumption being that the glyph is derived from $\text{ᯚ} k\bar{a}l^{1/4} \times \text{ᯚ} a\bar{r}a\bar{i}^{1/2}$. While such a “derivation” seems to go against the order of words in the name of this fraction: *arai-k-kāl*, and the analysis of the second half as being a deformed ᯚ requires some stretch of the imagination, in the end, the requested glyph isn’t really all that different from the existing one except that the finishing stroke to the right is higher and touches the cusp on the top in the proposed glyph.

However, in L2/12-231 p 9 we have recorded that this is the fraction with the most variants, and *none* of those variants show a contact between that stroke and the cusp. They

also do show the stroke below the baseline and not higher up as in the proposed glyph. As such, the existing glyph already seems most representative of all the variants and thus it is better not to change it to favour one particular variant.

§3.8. 11FCC ஈ௪- THREE SIXTEENTHS

On p 8, the GoTN requests that the glyph be changed to ஈ௪ to reflect the shape seen in inscriptions. I have already referred to the inadvisability of copying archaic epigraphic forms to the code charts since they are a set with other epigraphic forms of all the character of the script. The Unicode Standard 8.0 p 846 (ch 24.1) admonishes us that:

A representative glyph is not a prescriptive form of the character, but rather one that enables recognition of the intended character to a knowledgeable user and facilitates lookup of the character in the code charts.

Here the term “knowledgeable user” does not exclusively refer to epigraphists who would be the only people to recognize epigraphic written forms, but also to those who may be working with later period printed texts to digitize them. Such printed sources (for instance see pp 28 and 37 of L2/12-231) are replete with usages of the form ஈ௪- for this fraction. Since it is found in flowing text amongst other modern Tamil glyphs including other fractions, I feel ஈ௪- is necessarily the appropriate form to represent the character as part of the modern Tamil glyph repertoire seen in the code charts.

Now the GoTN doc says on p 6:

the glyph for 3/16th does not reflect the actual glyph; It is possible that the glyph is handwritten in the attestations to represent this glyph. It is not uncommon for various publications of old inscriptions to use handwritten or near-representative glyph, to denote a particular glyph.

This comment seems totally out of place since the aforementioned attestations produced in L2/12-231 from *printed* publications very clearly show ஈ௪- as a *printed* glyph and it is in fact the attestation provided by the GoTN for the epigraphic shape ஈ௪ that uses handwritten glyphs in a printed publication. The doc continues:

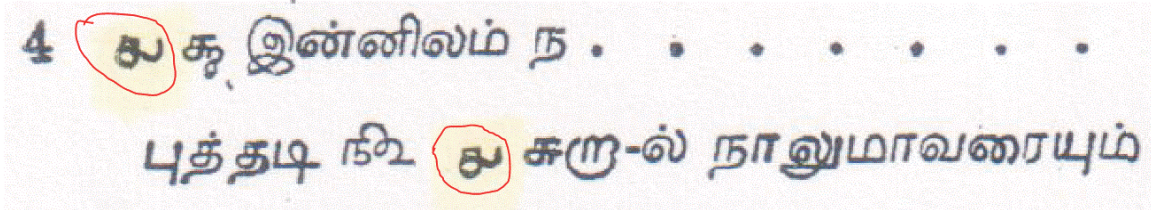
Unfortunately, such glyphs cannot be used for encoding and only those found in the original root documents are the right candidates.

Repeat here the earlier points about representative glyphs not being prescriptive and the incongruity of including an epigraphic glyph amongst modern ones.

While I agree that the older glyph ூ as a ligature better reflects its significance as three (௩) times one sixteenth (ௌ), I would still point out that it is the variant ூ- which has been used consistently across later texts and is thus more representative in the context of the modern Tamil script. The older glyph is just another archaic variant.

§3.9. 11FCD ூ ONE FIFTH

On p 22, the GoTN doc calls the current glyph “incorrect” – though it is attested by Beschi (L2/12-231 p 29) – and asks for it to be replaced by ூ, though this glyph is not sufficiently true to any of the attestations including that provided by the GoTN doc itself:



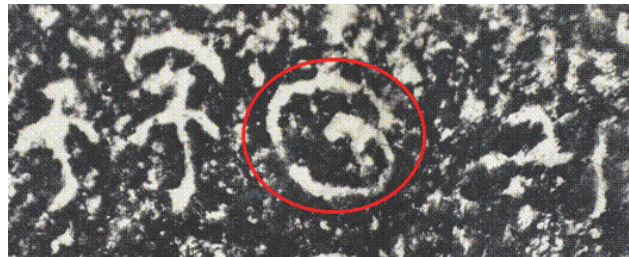
ூ is moreover confusable with 11FC4 ூ ONE FORTIETH. The point however is that the cup-like descending part to the right should be retained and in fact most of the attestations even in my original proposal (L2/12-231 p 38) indeed show this:








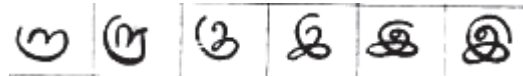
Thus the glyph should indeed be corrected but as seen above, not to ூ as suggested.



§3.10. ONE HALF - Additional character

The fraction one half is written as ூ and proposed for 11FCE. On p 18, the GoTN doc asks for separate encoding of the alternate representation of this fraction: ூ (seen in the original proposal as the somewhat but not overly different ூ). I have already discussed this issue in L2/12-231 p 9. I did not propose this as a separate character since I was (and am) not convinced that ூ/ூ is not just an earlier variant of the later simplified form ூ just as in the case of ூ vs ூ for one sixteenth above. The attestation provided:





... does show a long wrapped-around tail and a slight cusp in the middle as in , but the long tail could have been shortened and the cusp could well have reanalysed into a starting downward stroke and reconnect with the upward stroke on the left to form the loop of . The point is that as before in the case of  vs , it would be difficult to find contrastive usage, but since Caa Ganesan (L2/12-231 p 33) actually shows a different evolution of :





where the -like form later gets conflated with 0B87  TAMIL LETTER I, and later printings actually show the two alternatives side-by-side (L2/12-231 pp 35, 36):



If the UTC is satisfied with the distinctness of character identity, I shall separately submit a formal proposal (with character properties) for this alternate form of one half  along with  for one twentieth (see §1.1 and 1.5).

§4. Measures of grain




§4.1. 11FD1 னு NEL







The glyph currently proposed is based on Wickremasinghe (p 50 of L2/12-231) since the other sources often conflate it with 0B9C  TAMIL LETTER JA or a variant thereof. On p 26, the GoTN doc recommends replacing this with . Again, this seems to be an attempt to be “authentic” by giving “the original glyph from the attestation” but it forgets the fact that it is an archaic glyph found among other archaic glyphs for other Tamil characters. This is certainly a valid historical variant but not representative of the character for a code chart displaying modern Tamil glyphs.

§4.2. 11FD4 உரி URI


The GoTN doc p 48 notes that the proposed character is just a joint writing of the word உரி and hence recommends it not be encoded. Since I do not have (or expect to have) strong evidence of distinctive ligated usage, I have no objection to dropping this character.

§4.3. 11FD5 ஁ MUUVULAKKU

On p 33, the GoTN doc asks that the glyph be “modified a bit” to the form  to highlight the presence of ஁ denoting 3. However, the same doc has earlier on p 28 provided  as the glyph for this character as inferential support for  to be encoded for *ulakku* and

mentioned on p 30 that “logically”  is written combining  and . Clearly, this earlier comment supporting  for *ulakku* conflicts with the recommendation of  which does not show the initial curve for *muuvulakku*. In any case, this is a trivial difference in the shape of a ligature and the various attestations provided on pp 39-41 of the original proposal L2/12-231 support the existing glyph . As such the glyph is better left as is.

§4.4. 11FD6 PATAKKU




The GoTN doc p 32 notes that written forms such as  (which is only slightly different from the existing glyph) are found in the “primary resources”, that “research-worthy variants” are observed and finally that investigating the “true form” has been “a bit complex”. It asks that the character be postponed from encoding. It is unclear what defines a “true form” from epigraphy which changes from inscription to inscription and in any case, there is no doubt that the character exists and its current proposed glyph based on printed sources is also reasonably similar to earlier ones. Thus I see no need for postponing.

§5. Old Currency

§5.1. 11FD8 PAICAA

The GoTN doc on p 49 recommends that this character be removed from the ballot as there is insufficient evidence of consistent usage of this character. I had proposed this character mainly based on evidence from Winslow and Pope (L2/12-231 pp 45, 48) but had mentioned the inconsistency in glyph and name in my discussion (L2/12-231 p 14). As such, upon reflection I agree with the GoTN that this character is better rescinded for now.

§5.2. 11FD9 ANAA

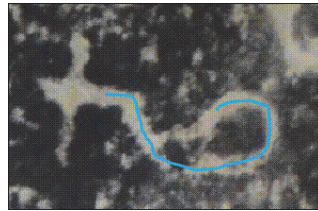
I had proposed this character based on evidence from Pope and Wickremasinghe (L2/12-231 pp 48-50). The GoTN doc on p 50 recommends that this character not be encoded as this seems to be merely a stylistic combination in archaic writing of the word *anā* viz . I have no evidence to specifically contrast between  and . Similar to the case discussed in §3.6, both these mean the same anyway and are unlikely to be available in contrast. As such, I have no objection to removing this character from the ballot.

§5.3. 11FDB ழ PANAM

On pp 33-35, the GoTN doc provides attestations for historical glyphs ழ and ழ for this character and asks to replace the current glyph which is based on modern attestations. I have sufficiently repeated the argument against using a historical glyph i.o. modern ones.

§5.4. 11FDC ே PON

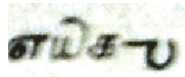
The current glyph was proposed for this character based on evidence from printings (L2/12-231 pp 45, 48, 49, 51). The GoTN doc on p 37 shows attestation for a shape similar to ே. While the actual attestation shows this ligated to a +like mark denoting 1:



... the glyph is indeed uniquely identified as the one for *kalañcu* which is a synonym for *pon* in the sense of gold coin of particular weight used as currency, as seen from the attestation on L2/12-231 p 21 (near bottom of page):



The publication transcribing the estampage also uses a similar glyph:



Curiously, the current glyph seen in generic printings ே seems to be almost a 180° inverted version of the earlier glyph seen in epigraphs and printings based on those sources. The evidence of usage of ே indicates that the form ே (attested only isolate in some references rather than actual usage) may be an inaccurate glyph substitution. Based on the evidence given, I support changing the glyph as recommended.

§5.5. 11FDD ௐ VARAAKAN

The current glyph was proposed based on attestations in printings (L2/12-231 pp 45, 48, 49, 51). While the GoTN doc on p 39 comments that this glyph seems to be “a bit far away from the actual”, the alternate glyph it proposes ௐ is not all that different, and seems to be a just a historical variant, since the difference is only that the initial ௐ part seems contracted and raised whereas the final ௐ part is lowered. Note that we already have pointed out in §1.2 that this is how the letter ௐ was written in those days. Given this, I

analyse the proposed alternate glyph as a historical variant. Thus I recommend retaining the current glyph for consistency with modern Tamil glyphs.

§6. Other symbols

§6.1. 11FDF கை KAJAM

On p 51 the GoTN doc recommends that this character “not be encoded as part of Tamil encoding” because the unit (leave alone its name) is not of Tamil origin but of Persian origin. However, this is not a criterion for not encoding a written character. The written form clearly belongs to the Tamil script, whatever be the cultural origins of what it denotes. Unicode encodes scripts, not languages or cultures. People should understand that the usage of a character in a script does not necessarily imply that what it signifies is native to a particular language or culture. As such, this character should certainly be encoded and specifically as a Tamil script character. There is hence no need to “take this symbol out of this proposal for further reason to identify the original source”, as the GoTN suggests.

§6.2. 11FE3 ஸூ DRY CULTIVATION

On pp 41-42, the GoTN doc recommends that the alternate glyph for this character viz ஸ_ௌ be chosen. Frankly, I had chosen the current glyph only for consistency with that for 11FE2 ஸ_ௌ WET CULTIVATION. I have no objections to changing it to the alternate attested form ஸ_ௌ.

§6.3. 11FE4 ஸ_ௌ LAND

On pp 42-45, the GoTN doc provides copious examples for an alternate glyph of this character where the lower part of the ஸ (na) loops around its body to join the starting stroke of ல (la) on the top right. I have no objection to this glyph change, except to caution that care should be taken to ensure that the final stroke is clearly curved up as in ல to differentiate from the rupee sign ரூ.

§6.4. 11FE7 ஸ_ௌ FIRST

As above, on p 46, the GoTN doc provides an alternate glyph for this character where the lower part of the first component glyph ஸ (mu) loops around its body to join the starting stroke of ல (la) on the top right to give something like ஸ_ௌ. Again, I have no objection to this glyph change so long as care is taken with the final stroke as above.

§6.5. 11FEA ீ CIRANCIVI

On p 53, the GoTN doc objects the the encoding of this character citing the example of one particular marriage invitation where the abbreviation was *not* used and the whole words were used instead. Certainly, the very notion of an abbreviation presumes the existence *and* previous/occasional usage of an unabbreviated form, so this is no argument against the encoding of this character which was proposed based on attestations (L2/12-231 pp 46, 48, 49, 50) and is thus per se a valid candidate for encoding.

The “first-letter” argument is not very strong either since it would be difficult to argue how this abbreviation is any less valid than the already encoded 0BF4 ீ MONTH or 0BF9 ீ (ீ) RUPEE. or the also-proposed and not objected-to 11FD0 ீ KIL, 11FDB ீ PANAM, 11FE0 ீ KULI, etc which are all derived from the first syllable of the word and thus could also arguably be replaced by that syllable.

However, I understand that there are cultural reasons to hesitate to encode this character which is anyhow arguably of less utility than the other “secular” characters. As such, I have no objection to removing this character from the ballot.

§6.6. 11FEB ீ PILLAI

On p 54, the GoTN doc recommends that this character not be encoded. Again, the real concern seems to be cultural and not technical. Again, since the character is probably of limited use, I have no objection to rescinding this character.

§6.7. 11FEC ீ RAAJA

On p 55, the GoTN doc recommends against encoding this character. Unlike in the above cases, there is no cultural objection to this particular term. Rather, the argument attempted is technical, that the symbol “can be derived by combining 0BF1 ீ ONE HUNDRED and 0BF3 ீ DAY. However, glyphic similarity or even identity of the visual components of an evident ligature to two unrelated characters is not sufficient cause to recommend its representation as a sequence (perhaps using ZWJ) of those two characters. While later printings have indeed used 0BF1 ீ ONE HUNDRED as an approximation due to absence of printer’s repertoire, this is not sufficient cause to reject the encoding of 11FEC ீ RAAJA. which is in fact a ligature probably based on the old spelling of the syllable *rā* viz ீ as I have explained in L2/12-231 p 17. As such, this character should be encoded.

§6.8. 11FED ௩ KKU

The discussion in §3.6.9 (pp 55-59) of the GoTN doc seems to imply that this character should be rejected as it is merely a ligature. However, this disregards the detailed argument which had been presented earlier in the original proposal. To summarize, while linguistically and orthographically, this sign is indeed also a ligature of *kku*:



... it is used contrastively to the other ligature in that the ligature ௩ is used specifically in numeric contexts as discussed in detail in L2/12-231 p 20 as against the normal linguistic usage of the ligature ௩ as ably demonstrated by the GoTN doc on p 58.

As such, while ௩ should be analysed and encoded as a ligature, ௩ should be encoded atomically due to its unique and contrastive usage. See also §7.5.

§6.9. 11FEE யு YUM and 11FEF வு VUM

Based on the evidence for the existence of very many more ligatures provided in §3.6.9 (pp 55-59) of the GoTN doc (some of which had briefly been hinted upon in L2/12-231 p 19) I agree that these two characters should not be encoded and instead treated as ligatures to be supported by a font appropriate for the provenance of the document to be imitated.

§7. Normative character names

§7.1. Usage of Tamil even when translatable?

On pp 59 and 60, the GoTN doc recommends that fractions should be named per their Tamil names. However, the Unicode Standard is intended for international use and normative names are primarily defined in the English language. As per ISO 10646* Annex L p 2428 (p 106 in main PDF) *Character Naming Guidelines*, “the names of characters denote their customary meaning” (in English). It is only when words (such as ANUSVARA) are not meaningfully translatable into English that one uses native words written in limited A-Z Latin script. As such, it would be appropriate to retain the existing proposal of naming the fractions in English and providing them readable annotations in Tamil.

Even apart from the fractions, the GoTN has proposed the usage of Tamil words as part of normative character names in the case of the following characters though their

* http://standards.iso.org/ittf/PubliclyAvailableStandards/c063182_ISO_IEC_10646_2014.zip

names are quite translatable: agricultural symbols 11FE2 to 11FE5, 11FE6 CHARGE, 11FE7 FIRST, and additional clerical signs 11FF0-11FFF. This also goes against the above guidelines. For the “recognizability by natives” requirement we have already provided appropriate annotations. As such I recommend retaining English names as normative in all these cases as these can be understood by a much wider audience.

§7.2. 11FE6 ஷ CHARGE and 11FE7 ஸ் FIRST

Now I agree with the GoTN that the English renderings of the Tamil words *vacam* and *mutal* in the names of 11FE6 CHARGE, 11FE7 FIRST can be improved upon, but so improving is the solution instead of totally switching language. I recommend that 11FE6 be renamed as TAMIL IN POSSESSION SIGN and 11FE7 as TAMIL STARTING FROM SIGN.

§7.3. Spelling in usage of Tamil when untranslatable

As mentioned in my original proposal L2/12-231 p 61 when the native words are not translatable into English, I have used Tamil words as part of normative character names. The matter of the romanization of these words was given much importance by certain parties in 2014 and as a result, an expert committee was appointed and met twice to decide upon a scheme (ref L2/15-060R). Based on further feedback this scheme was slightly modified (ref L2/15-079) and this is the scheme currently reflected in the latest draft of ISO 10646 5th ed CD2 L2/15-339 N4705.

Now the later appointed expert committee which has produced the feedback documents L2/16-039 and the superseding L2/16-062 does not seem to have given much attention to the spelling issue and no de jure spelling scheme is prescribed by them. However, the spellings used on pp 60, 61 of L2/16-062 thankfully largely follow L2/15-079 and hence L2/15-339 with a few changes: not differentiating short/long E/O, using TH for த rather than T, using ZH for ஸ் rather than L. This leads to the following de facto scheme:

அ	ஆ	இ	ஈ	உ	ஊ	எ ஏ	ஐ	ஒ ஓ	ஔ
A	AA	I	II	U	UU	E	AI	O	AU

க	ச	ட	த	ப	ந் ன் ண் ற் ன்	ம
K	C	T	TH	P	N	M

ய	ர ற	ல ள	வ	ஸ்
Y	R	L	V	ZH

However, even this scheme is not consistently followed in the proposed names. While AAVANAMUTRU for 11FFF instead of AAVANAMURRU can be seen as a nod to pragmatism (since it's actually pronounced similar to TR and nor RR), the names 11FD2 CHEVIDU and 11FF3 CHILLARAI employ out-of-pattern CH for C and D for T. I have however recommended that 11FF3 and 11FFF (among others) should retain their English names. 11FD2 CHEVIDU is however one of the characters where the native name is untranslatable. I discuss it below.

§7.4. 11FD2 சுவடு CUVATU

On p 27, the GoTN doc notes that “the name of this measure Suvadū is varying over an extended period and regions [sic]” but claims that “the right name of this measure is Chevidu” among all those varying names! No linguistic or other bases for this claim are provided. While attestation for the usage of the word “Chevidu” is provided, the question is not one of attestation but one of linguistic primacy. The linguistic evidence points to *cuvatu*¹ being the primary form and not *ceviṭu*² (rendered as *cevūtū*).

In Tamil, lenition in colloquial usage from *uva-* to *evū-* is seen as in *cubar* (wall) to *cevūrū*, but not the other way around. Similar lenition from *uva* to *ō* as in the case of *tuvaramparuppu* (pigeon pea) to *tōramparuppu* is seen and can explain the alternate rendering *cōṭu*³ as well, whereas taking *ceviṭu* as the primary form cannot explain that. As such, it is most curious that the GoTN committee recommends *ceviṭu* as the primary form in contradiction to this linguistic evidence.

Perhaps the committee feels that *ceviṭu* is more often seen in inscriptions (judging by their preference for forms seen in such contexts) but one should also remember that inscriptions often use quite freely what are grammatically considered colloquialisms such *kuṭuttāṇ* for *koṭuttān*. As such, they may be proof of historical written forms but they are not proof of originality or linguistic authenticity of a word as compared to another.

Based on the above evidence, I recommend retaining the spelling CUVATU as is now.

§7.5. 11FED சுவடு KKV

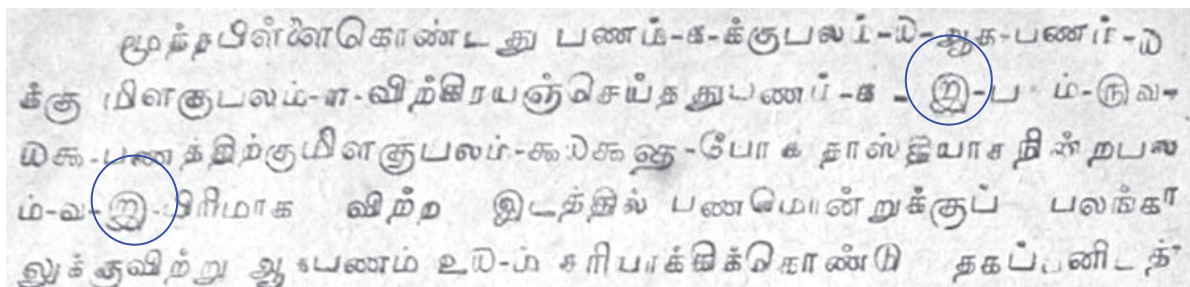
In the background of the discussion in §6.8, I feel that this character should be renamed to TAMIL FOR SIGN since that explains its usage much better. KKV refers to its linguistic content

¹ <http://dsalrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.5:1:1400.tamillex.537331>

² <http://dsalrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.5:1:3479.tamillex.1374072>

³ <http://dsalrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.5:1:4666.tamillex.1843782>

but this character is only ever used in the particular sense of “for <number> <items>”. For instance, the relevant portion of the text below translates to “5¼ palam-s for 1 paṇam ... for the remaining ¼ palam, ...”:



Thus I request that 11FED TAMIL ABBREVIATION KKU be renamed to TAMIL FOR SIGN. It should retain its existing annotation to clarify the usage context.

§7.6. Disambiguating characters with identical connotations

The GoTN seems to recommend -1 -2 as suffix to disambiguate different characters with identical connotations. I do not know if the UTC will recommend the disunification of the two forms of ONE SIXTEENTH and ONE HALF. (See §3.6 and §3.10.)

Even if these two fractions are disunified and thus disambiguated in naming by suffixing -1 and -2 (since there is no better way), I feel the names of 11FF1 TRADITIONAL NUMBER SIGN and 11FF2 TRADITIONAL CREDIT SIGN should not be changed that they carry the additional connotation of earlier usage. The fact that the names of the existing 0BFA NUMBER SIGN and 0BF7 CREDIT SIGN cannot be changed to suffix -1 should also be considered.

§7.7. Other fine naming issues

One also notes that the GoTN seems to desire the use of the word SYMBOL. I would point out that the word SYMBOL has more graphical connotations whereas SIGN is more inclusive to allow linguistic content in combination with graphical ones such as in “exit sign”, “keep left sign” and so on. The word SIGN is also thrice as common as the word SYMBOL in the Unicode standard. To minimize changes, the word SIGN may be retained.

Further, the GoTN recommends differentiating the various categories of signs: MEASURE, CURRENCY, WEIGHT, AREA, LAND, COMMON, CLERICAL, ACCOUNTING etc. This seems to be overkill and the common usage of the characters are already clarified by the subheadings in the names list. I feel it is better to keep the names as brief as possible.

§8. Annotations

§8.1. Tamil annotations

On p 60, the GoTN doc recommends that “the names are also written in Tamil as annotation”. Full-Unicode annotations for the charts is probably a desideratum not only for Tamil but also for other scripts/languages. This is currently not implemented, however.

§8.2. Corrections to values etc

As regards specific annotations, the GoTN has (on pp 35, 36, 39, 72) some comments on the specific values of the currencies/measures denoted by some signs. One obviously has to expect a certain amount of regional/temporary variation in such historical usage. Since I expect other editorial changes as a result of the addition and/or deletion of characters due to the GoTN’s feedback, I defer my giving specific input regarding annotations until we have a clear picture of the final character set to be encoded based on which I can provide a comprehensive set of changes for the names list.

§8.3. 11FE5 அல SALT PAN

On p 61, I note that the GoTN has changed the Tamil rendering of 11FE5 அல to *uppaḷam* instead of just *aḷam*. However, the Tamil Lexicon clearly indicates* that the word *aḷam* is sufficient to indicate the sense of a salt pan. Furthermore, the graphical derivation of the sign அல is clearly from அளம் *aḷam* and not from உப்பளம் *uppaḷam*. In fact, the very usage *uppaḷam* only serves to clarify the sense of the usage of *aḷam* as a place of production of salt and not generic seaside/salty/brackish land. Thus the Tamil rendering of the word should remain *aḷam* true to the linguistic derivation of this sign. A native reader reading “*aḷam*” after having read “salt pan” will easily remember the specific word *uppaḷam*.

As mentioned above, I shall provide further editorial input re the names list after the UTC decides on which characters to encode so that the character set can be consolidated and the codepoints finalized.

-o-o-o-

* <http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.0:1:4649.tamillex.1717107>