Request to deprecate 0B82 TAMIL SIGN ANUSVARA

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

2011-Dec-14

In the Tamil Unicode encoding there is the character 0B82 TAMIL SIGN ANUSVARA:

0B82   ⚬   TAMIL SIGN ANUSVARA
• not used in Tamil

This document requests to deprecate the above character and annotating it recommending that it is not to be used. as if it is misused it can lead to collation, searching and other text processes failing.

Misleading nature of the encoded character

It is a well known fact that Tamil language phonology does not use the anusvāra which is basically a Sanskrit-based sound and written form. Thus the Tamil script does not natively have a written form for the anusvāra. As such, why this character was encoded in the first place is unknown.

The Unicode code chart seeks to address the situation by giving an annotation "not used in Tamil". Unfortunately, this annotation is not clear enough. People have interpreted it to mean that this is not used for the Tamil language and it may hence be used for Sanskrit language words written in Tamil. For example, see the following passage from a 2007 masters dissertation on Tamil typography by one Fernando Mello (retrieved 2011-Jan-24 from http://www.leonidas.org/rdg/matd/dissertation/FernandoMello_dissertation.pdf):

13 Despite the fact that the pulli is usually represented by a dot ('), some typefaces use an outlined circle (˚) for the sign. G.U. Pope says in *A Tamil Handbook* that the pulli can be represented by a dot or an outlined circle (p. 8). A detail from the *Tamil Lexicon* shows different typefaces used together in a page having the pulli expressed by a circle or a by dot (figure 7). However, the actual Unicode standard chart for Tamil reserves characters for a dot, which is described as 'Tamil sign Virama' (uni0BCD) and for an outlined circle, which is described as 'Tamil sign Anuswara' (uni0B82). The Anuswara, as in other indic scripts, may be used to transliterate Sanskrit loanwords in Tamil, but is obsolete and not used in modern writing. Above-base mark OpenType features under the names 'Virama' and 'Anuswara' for placing these dot or circle characters above consonants can also be found in contemporary Tamil types.

Virama

Anuswara

The above text also indicates that Tamil fonts mark their "anusvara" glyphs by the OpenType feature "anusvara". This is meant for the anusvara glyphs of scripts which have genuine anusvara-s and to apply it to the non-existent Tamil anusvara is meaningless.

Furthermore, the IndicSyllabicCategory.txt also contains the line:

```
0B82          ; Bindu # Mn      TAMIL SIGN ANUSVARA
```

which is again totally misleading to implementers indicating that this character is actually used in the Tamil script for anything at all. While Sanskrit-based words are also written in the Tamil script, and even entire Sanskrit texts have been published in the Tamil script with appropriate orthographic notations, this "Tamil anusvara" is not attested even there!

True status as a glyphic variant

Now it may be suggested that while this character so far is not attested, it may be used for something new in the Tamil script. However, this is not possible. This is because, the written form of a circular ring above consonants is already attested in Tamil, only not as a separate character but as in fact a glyphic variant of the Tamil pulli virama.

The "pulli" is a dot-like or small circle/ring-like glyph which is used as the Tamil virama and historically in the Tamil script for marking the short vowels E/O. Three pulli-s triangularly placed together make the Tamil āytam. It is clearly seen in writings and printings of the Tamil script that in all three uses – as the virama, for the short E/O and in the āytam – the pulli may be either a dot or a small circle/ring.

Tamil Virama	:	க்	க̊
Tamil Short E/O (historic)	:	ெ̇க/ெ̇கா	ெ̊க/ெ̊கா
Tamil Āytam	:	∴	° °

I have attestations for all three cases of glyphic variation and can provide separately if requested. In this document I only show attestations in the case of the virama.

The following samples from Kaṇakkatikāram, a mathematical treatise published in the year Prabhava (1868 or 1928) (http://www.tamilcc.org/thamizham/ebooks/3/273/273.pdf) shows the circular pulli used as a virama:

The texts read:

> *tiruccirrampalam*
>
> *kaṇakkatikāram*
>
> *kaṭavuḷ vaṇakkam – civattaip pōrral*
>
> *venrulē pulan kaḷaintār meyyuṇaruḷḷantōruñ*
>
> *cenrulēyamutamūrrun tiruvaruḷpōrriyērrik*
>
> *kunrulēyiruntu kāṭci koṭuttaruḷ kōlam pōrri*
>
> *manrulē māriyāṭu maraiccilampaṭikaḷ pōrri*

One notices that there is no anusvāra in the above passage, as indeed there is no anusvāra in Tamil! Thus the circular pulli is evidently used for the virāma and virāma alone.

Of course, there is no question that the dot is also used as the pulli for the virama, as currently shown as representative glyph in the code chart. A sample from p 51 of Tolkāppiyam Eḻuttatikāram, 1937, C Ganesaiyer, Cunnakam, Sri Lanka:

Conclusion

It is thus clear that the pulli in Tamil, specifically when used as a virama, is presented both as a dot and a small circle/ring.

The representative glyph of the Tamil virama in Unicode is shown as a dot:

0BCD   ○̇   TAMIL SIGN VIRAMA

While this is not per se a problem, the contrastive presence of the so-called Tamil anusvara:

0B82   ○̊   TAMIL SIGN ANUSVARA

... causes readers (such as the author of the masters thesis quoted before) to think that the anusvara is circular and the virama is a dot. The fact is however that there is no anusvara and the virama can be presented as both a circle and a dot. This is the nature of the Tamil script. Thus it is clear that the encoding of 0B82 TAMIL SIGN ANUSVARA was a mistake.

While native users of the Tamil script will immediately identify this mis-encoding for what it is, non-native users (such as the author of the masters thesis quoted before) will be misled by what is given in the Unicode code chart.

The presence of this character not only misrepresents the usage of the pulli in the Tamil script, but it also causes unnecessary presence of an "anusvara" OpenType feature in Tamil fonts, and an unnecessary entry under "Bindu" in the IndicSyllabicCategory.txt.

While these may be considered mere useless appendages, the real problem arises when this character is misused to represent the circular glyphic variant of the pulli. While one cannot prevent legitimate distinct characters like 210A SCRIPT SMALL G etc being misused as glyphic variants, the Tamil anusvara is not even a legitimate character. If it is misused as a glyphic variant of the Tamil virama, then it is clear that text processes such as search, collation etc will fail, because the two characters are not canonically equivalent.

Noting the mis-encoding of this character, the Tamil Nadu Government circular on the use of Unicode for Tamil says (p 7 of L2/10-318):

> *The character Tamil Sign Anusvara (U+0B82) shall be treated as an invalid character in official documents.*

However it is not sufficient that the character be prohibited from TN Govt official documents but in all Unicode text in general as indicated above.

Action to be taken

Since the presence of this character is thus not only meaningless but also problematic, it is requested that the character be deprecated. This deprecation must be noted in the Unicode chapter on Tamil and also in the code chart. The recommended annotation to replace the existing one is as follows:

- character encoding is a mistake
- usage strongly discouraged
- use 0BCD TAMIL SIGN VIRAMA instead

The meaning of the phrase "strongly discouraged" is as in TUS 6.0 §3.4 D13. If possible, an outright wording of "do not use" is preferred.

In relation to this, it is also requested that 0BCD TAMIL SIGN VIRAMA be given an informative alias as "pulli", just like 094D DEVANAGARI SIGN VIRAMA and 09CD BENGALI SIGN VIRAMA have the informative aliases "halant" and "hasant" respectively.

An annotation to the Tamil Virama to the effect:

- may be presented as a dot or a ring

is also recommended.

–o-o-o–