

## Using superscripts and subscripts with Tamil consonants

---

*Muthu Nedumaran*

This document is written in response to a proposal submitted to the UTC By Shriramana Sharma, (L2/10-256R) to encode characters for a newly coined term "Extended Tamil".

While others have written with linguistic and grammatical arguments on why this proposal should be rejected, this document addresses just one point the author mentions in his response to INFITT's objection:

*<Quote>The existing Unicode model does not support the proper desired rendering as it would force one to place the superscript digits after any vowel signs, and that is not right, as it is neither the place nor intention of Unicode to force orthographic reform. Unicode is to support to the best of its abilities the existing usage as evidenced by proper attestations. Such existing usage was shown in my proposal L2/10-256R as also in the follow-up document. To enable Unicode to support such existing usage and solve the rendering problem (among other matters), these characters were proposed and should be encoded.*

*Having not even hinted that they are interested in or concerned about addressing the rendering problem described in my previous documents as above, the INFITT people have brushed the matter off (in their email on the unicode list dated 2010-Sep-30) saying:*

*Tamil letters with superscripts ... for almost 10 years ... are encoded as sequences in Unicode.*

*Without properly addressing the rendering problem stated above, any party claiming to be interested in the proper representation of the Tamil script in Unicode should not merely decry the proposed encoding with passive statements like the above.</Quote>*

First, we understand that the common practice is to place the superscripts or subscripts 'after' the vowel-sign. However, for some reason, if the author wants it placed as desired, it should not warrant a change in the standard or introduction of new code-points.

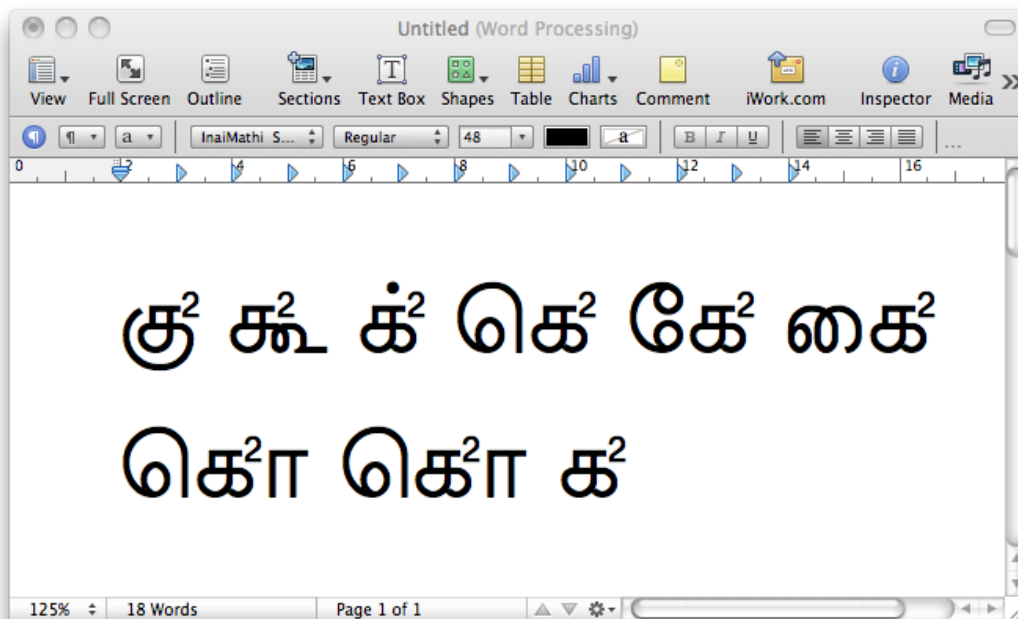
We need to decouple the issue of encoding from the issue of rendering.

The screen-shot below shows the rendering of the text as desired by the author in the Mac OS X, Pages application.

In the back-end store, the text is represented appropriately as

[consonant][superscript | subscript][vowel-sign]

With the required rendering rules defined in AAT, the superscripts or subscripts are placed next to the base consonant and not after the vowel-signs. This works for both left, right and split vowel-signs.



The above may not be possible with OpenType on Windows. It does not help even if the [consonant][superscript | subscript] pair is defined as akhands. This should be looked at as an implementation issue with OpenType rather than a standards issue with Unicode.

We had a similar problem in WindowsXP where the Aytham (U+0B83) was wrongly implemented as a sign and appeared with a dotted circle when placed before a consonant. The implementation later fixed this 'bug'.

When the standard allows the use of U+00B2, U+00B3 and U+2074 along side consonant characters to preserve a certain distinction, the implementations should follow. The standard should not be modified to fix an implementation bug.

### Word boundaries

The author has a point about issues with word-boundaries. U+00B2, U+00B3 and U+2074 are GC=No.

The general understanding is that all Sanskrit text processing will be done in Devanagari and only the final output is rendered in the transliteration-target script.

If combining characters are absolutely required for the sake of word boundaries, then they are the only ones that may need to be encoded:

COMBINING SUPERSCRIPT TWO

COMBINING SUPERSCRIPT THREE

COMBINING SUPERSCRIPT FOUR

The above are just the three that can replace U+00B2, U+00B3 and U+2074.

There are many examples of text written with subscripts as well. These combining characters may also be applicable in other scripts. The details can be in a new proposal.

In conclusion, it may be best to identify and encode the required combining characters with the appropriate character properties rather than encoding 'pre-composed glyphs' as Tamil letters.

*Muthu Nedumaran*

*Founder & CEO, Murasu Systems Sdn. Bhd., Malaysia*

*Founder Vice-Chair and Former Chair, INFITT*

*muthu(at)murasu(dot)com*