

Proposal to Annotate Brahmi Sign Anusvara

Vinodh Rajan vinodh@virtualvinodh.com

Shriramana Sharma jamadagni@gmail.com

1 Introduction

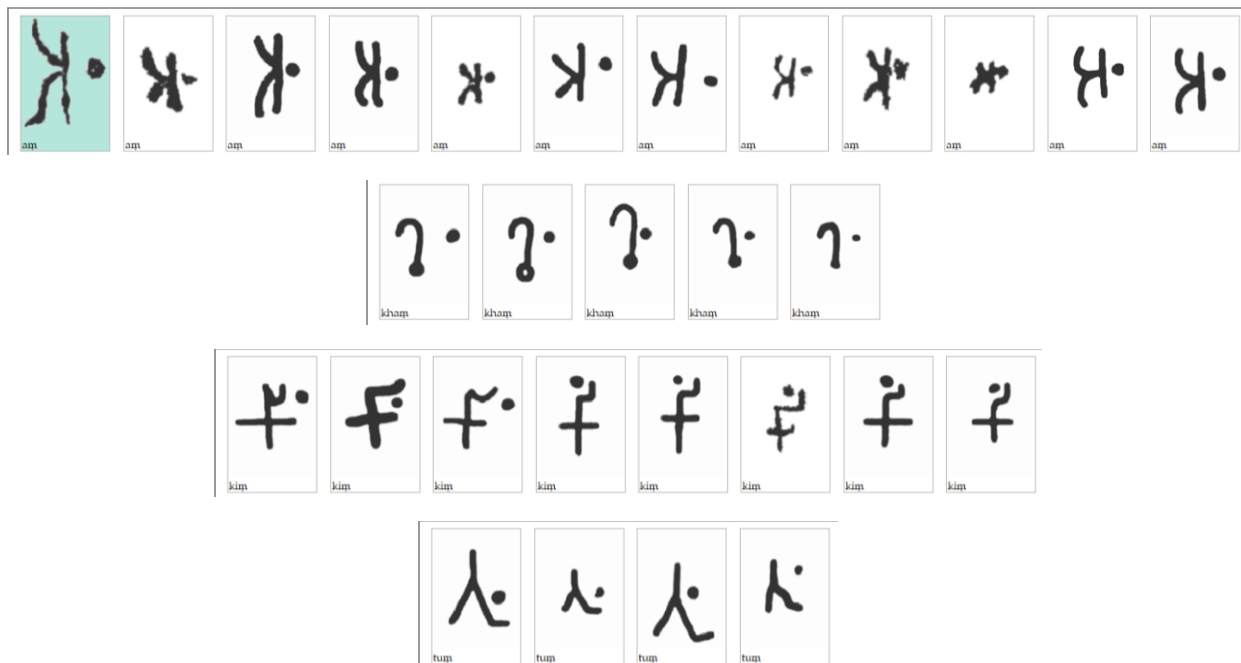
L2/19-402 initially proposed the addition of 6 Old Tamil characters to the Brahmi block, out of which 5 were accepted by the UTC #162 in January 2020. The dot-shaped Old Tamil Virama was not accepted due to the block already having another dot-shaped character i.e. U+11001 the Brahmi Anusvara character, which could possibly be repurposed to represent the Old Tamil Virama. It was also suggested to reconsider the unification with the existing Virama character.

This document proposes the unification of the Brahmi Anusvara and the Old Tamil Virama and requests that the existing Brahmi Anusvara character be annotated for its additional usage as the Tamil-Brahmi Virama.

2 Positioning of Brahmi Anusvara and Tamil-Brahmi Virama

The Brahmi Anusvara is a dot-shaped character usually placed at the top-right position. However, it may also be placed at the top or even at the bottom-right position.

Below shown are examples from *Indoskript*¹ for occurrences of the following syllables - /am/, /kham/, /kim/, /tum/ & /lum/ - between 300 BCE and 100 CE.



¹ <http://userpage.fu-berlin.de/falk/>



Old Tamil Virama is also a dot-shaped character, which can occur in multiple positions.



iva kuṇarātū uṛaiyuḷa pātanatāṇa eri āritāṇa | atatuṇvāyi araṭṭa kāyipaṇa |

From *Early Tamil Epigraphy* by Iravatham Mahadevan

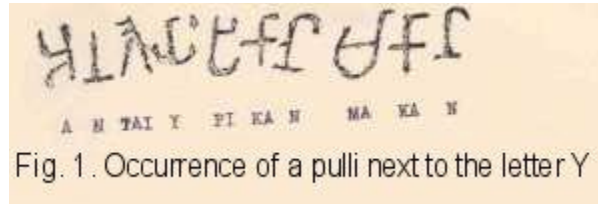


Fig. 1. Occurrence of a pulli next to the letter Y

https://www.cmi.ac.in/gift/Epigraphy/epig_vikramangalam_not%20considered.htm

Standardized representation of Tamil in recent books also shows variation in positioning:

	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஓ	ஔ	ஆய்தம்
	𑌀	𑌁	𑌂	𑌃	𑌄	𑌅	𑌆	𑌇	𑌈	𑌉	𑌊	ஃ
க	𑌋	𑌌	𑌍	𑌎	𑌏	𑌐	𑌑	𑌒	𑌓	𑌔	𑌕	𑌖
ங	𑌗	𑌘	𑌙	𑌚	𑌛	𑌜	𑌝	𑌞	𑌟	𑌠	𑌡	𑌢
ச	𑌣	𑌤	𑌥	𑌦	𑌧	𑌨	𑌩	𑌪	𑌫	𑌬	𑌭	𑌮
ஊ	𑌯	𑌰	𑌱	𑌲	𑌳	𑌴	𑌵	𑌶	𑌷	𑌸	𑌹	𑌺
ல	𑌻	𑌼	𑌽	𑌾	𑌿	𑍀	𑍁	𑍂	𑍃	𑍄	𑍅	𑍆
ண	𑍇	𑍈	𑍉	𑍊	𑍋	𑍌	𑍍	𑍎	𑍏	𑍐	𑍑	𑍒
த	𑍓	𑍔	𑍕	𑍖	𑍗	𑍘	𑍙	𑍚	𑍛	𑍜	𑍝	𑍞
ந	𑍟	𑍠	𑍡	𑍢	𑍣	𑍤	𑍥	𑍦	𑍧	𑍨	𑍩	𑍪
ப	𑍫	𑍬	𑍭	𑍮	𑍯	𑍰	𑍱	𑍲	𑍳	𑍴	𑍵	𑍶
ம	𑍷	𑍸	𑍹	𑍺	𑍻	𑍼	𑍽	𑍾	𑍿	𑎀	𑎁	𑎂
ய	𑎃	𑎄	𑎅	𑎆	𑎇	𑎈	𑎉	𑎊	𑎋	𑎌	𑎍	𑎎
ர	𑎏	𑎐	𑎑	𑎒	𑎓	𑎔	𑎕	𑎖	𑎗	𑎘	𑎙	𑎚
ல	𑎛	𑎜	𑎝	𑎞	𑎟	𑎠	𑎡	𑎢	𑎣	𑎤	𑎥	𑎦
வ	𑎧	𑎨	𑎩	𑎪	𑎫	𑎬	𑎭	𑎮	𑎯	𑎰	𑎱	𑎲
ழ	𑎳	𑎴	𑎵	𑎶	𑎷	𑎸	𑎹	𑎺	𑎻	𑎼	𑎽	𑎾
ள	𑎿	𑏀	𑏁	𑏂	𑏃	𑏄	𑏅	𑏆	𑏇	𑏈	𑏉	𑏊
ற	𑏋	𑏌	𑏍	𑏎	𑏏	𑏐	𑏑	𑏒	𑏓	𑏔	𑏕	𑏖
ன	𑏗	𑏘	𑏙	𑏚	𑏛	𑏜	𑏝	𑏞	𑏟	𑏠	𑏡	𑏢

http://know-your-heritage.blogspot.com/p/blog-page_14.html

Old Tamil Virama occurring at the top position

3 Unification of Old Tamil Virama & Brahmi Virama

This is clearly a non-starter; it just wouldn't allow a font to support both Tamil-Brahmi and generic Brahmi orthographies. With minority orthographies, it makes sense to remove font-level dependencies. With such unification, by default, a text would appear as generic Brahmi with the OS/Application always switching to the default system font (which is probably designed to capture the majority context). Tamil-Brahmi text will always appear as generic Brahmi text in Social Media etc. unless a specific font supporting Tamil-Brahmi is chosen by users.

Apart from the difference in shape between the Tamil/generic Brahmi Viramas, the Virama further causes conjunct/ligature formation between consonants which is not desired in Tamil-Brahmi. Hence Tamil-Brahmi Virama cannot be unified with the existing Brahmi Virama character which caters to generic Brahmi.

4 Unification of Old Tamil Virama & Brahmi Sign Anusvara

Anusvara does not occur in Old Tamil and as shown above both Anusvara and the Old Tamil Virama share the same shape and positioning variation. Any chosen positioning of U+11001 can be considered valid for its role as both the Brahmi Anusvara and the Old Tamil Virama. U+11001 also has the additional advantage of marking the end of an orthographic syllable, because in Tamil orthography, the Virama always marks the end of an orthographic syllable whereas this is not true in non-Tamil orthography.

Thus, Brahmi Sign Anusvara U+11001 can ideally be repurposed to represent the Tamil Brahmi Virama as well. Such unification of similar looking dots/circles already exist in TUS. The unification of the gemination marker and the spacing anusvara of Grantha (L2/13-061) may be cited as an example.

Since fonts and rendering engines already support U+11001, and its default shape/positioning in the two existing OS vendor fonts - Noto Sans Brahmi & Segoe UI Historic is also valid for the Old Tamil Virama, it has the further advantage of removing additional implementation overheads.

5 Alternate Positioning

Given that the unified Anusvara/Old Tamil Virama can occur in multiple positions, it would be useful if there would be some mechanism to fine-tune its positioning on a per character basis. The authors would like to receive feedback from the UTC about the ideal way to represent the various alternate positionings of U+11001.

6 Proposal

It is proposed that the following annotation be added to the Brahmi Code Chart for U+11001.

- *also used for representing the virama/vowel canceler of Tamil-Brahmi*

The Brahmi chapter of TUS must be modified to document this additional function of U+11001.

7 Tailored Collation for Old Tamil Brahmi

Since this character when used for Tamil marks the vowel canceler (*puḷḷi* in Tamil), its collation differs from when it is used for the Anusvara outside Tamil. This requires that a language-based tailored collation be applied to Brahmi text.

In Tamil-Brahmi, CONSONANT with U+11001 would be collated before a consonant with inherent vowel or vowel sign, just like in modern Tamil:

$$C + 11001 < C + 11001 + C2 < C < C + C2 < C + VS < C + VS + C2$$

In generic Brahmi, CONSONANT with U+11001 means a consonant with inherent vowel followed by the Anusvara and, hence, must be sorted immediately after the bare consonant with inherent vowel but before the same followed by any other characters, whether other consonants or vowel signs:

$$C < C + 11001 < C + 11001 + C2 < C + C2 < C + VS < C + VS + C2$$