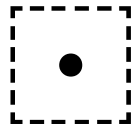# Proposal to add two characters for Brahmi

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

2012-Jun-12

§1. Characters to be encoded

11070 BRAHMI SIGN PULLI

11071 BRAHMI LETTER OLD TAMIL LLA

§2. Background and attestation for Tamil Brahmi Pulli

The Brahmi proposal L2/07-342 had noted (on pp 7-8) that during the early adaptation of the Brahmi script in the Tamil region (Early Tamil Brahmi) during the 2nd cy BCE to the 1st cy CE, there was no vowel-canceler in use, and:

1) In one style known as TB-I, the nominal form of the consonant denoted the pure vowelless consonant, with the consonant with the vowel sign AA denoting the consonant with either short A or long AA.

2) In another style known as TB-II, the nominal form of the consonant denoted both the pure consonant and the consonant with the short vowel A, with the consonant with the vowel sign AA denoting only the consonant with the long AA.

Due to the ambiguity in both above systems and due to pressure to conform with the Brahmi of neighbouring regions, a further orthographic modification later on (Later Tamil Brahmi) during the 2nd to 4th centuries CE led to a style known as TB-III where:

a) The nominal form of the consonant denoted the consonant with the short vowel A, making this vowel inherent as in other varieties of Brahmi,

b) The consonant with the vowel sign AA denoted only the consonant with the long vowel AA.

c) The pure vowelless consonant was denoted by the nominal form of the consonant with a puḷḷi i.e. dot acting as vowel-canceler.

1

The earliest attestation of the pulli in Tamil Brahmi as a vowel-canceler is from the 1st century CE, in a coin of one Vāsiṣṭhīputra Puḻumāvi (from ref 1 p 200):

aracaṉaku vāciṭṭi makaṉaku tiru puḻumāviku

The red arrow indicates the usage of the pulli and the transcription has the letter in bold.

Another early attestation is the Āṉaimalai inscription of the 2nd century CE: (ibid, p 402)

*iva kuṉaratū uṛaiyuḻa pātanatāṉa eri āritaṉa | atatuvāyi araṭṭa kāyipaṉa |*

Now the pulli in old Tamil writing was also used to mark short E/O as per the Tolkappiyam (early Tamil grammar) stating: "meyyiṉ iyaṟkai puḷḷiyoṭu nilaiyal, ekara-v-okaratt-iyaṟkaiyum aṟṟē" (ref 3 pp 51-52) meaning "The nature of the (written form of a) vowelless consonant is to stand with a pulli, and the nature of ĕ and ŏ is likewise".

An early example is from the Kuṭumiyāmalai inscription of the 3rd century CE (ibid, p 428) which also shows it being used as a vowel-canceler:

nāḻaḷ kŏraṟantaya pa?ya

Another example (for E) is from the Araccalūr inscription of the 4th century CE (ibid p 440):
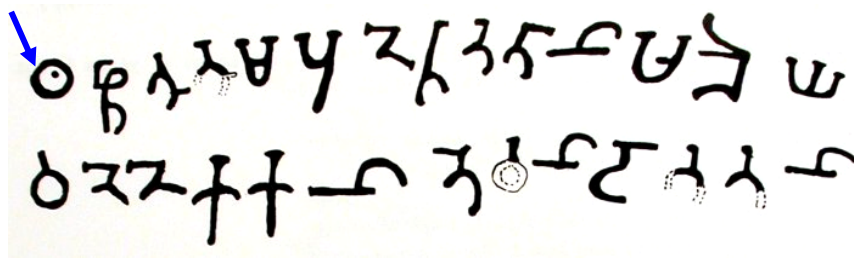
ĕḻutatuma puṇarutatāṉa malaiya | vaṉaṉakakaṉa tevaṉa cātataṉa ‖

The fact is that the Brahmi script reached the Tamil speaking region very early at a time when the northern version of the script had *not* developed a vowel-canceler. The Tamils clearly independently developed the pulli as a vowel-canceler*, and later extended it to shorten the vowels E/O also.

As regards the virama of Northern Brahmi, Salomon says (ref 2, p 37), under the heading of "The first through third centuries AD":

> *Among the new developments in this period, several characters such as …*
> *halanta (vowelless consonants) first came into common use.*

This virama of Northern Brahmi takes the shape of a horizontal stroke above:

… which written form is never seen in Tamil Brahmi. Prior to or parallel to the development of the Northern Brahmi virama, the Tamils had independently developed the pulli as a vowel canceler/shortener. It hence deserves to be recognized on its own rather than just as a glyphic variant of the northern virama.

---

* It is recognized that the identification of the language of the first-provided example. the coin of Puḻumāvi, is disputed as to whether it is Tamil or Telugu (ref 1, pp 199, 201). However, there is no doubt that the language of the later Brahmi writings using the pulli is Tamil. It is also only Tamil grammars that document the usage of the pulli for these purposes (ibid, p 230, §6.9). Thus the pulli in orthography is clearly a Tamil innovation.

§3. Justification for disunifying Tamil Brahmi Pulli

§3.1. Disunification from the Northern virama

The Brahmi proposal L2/07-342 states on p 8:

> *The two functions of Late Tamil Brāhmī puḷḷi can be subsumed under the heading of 'vowel reduction' (short to zero, and long to short), and puḷḷi should be encoded as 11046 BRAHMI SIGN VIRAMA; the Brāhmī virāma character can thus follow both consonant characters and the vowel characters e and o, in contrast to the modern scripts' virāma characters.*

However, as shown above, the Tamil Brahmi pulli was an entirely independent innovation. Furthermore, its behaviour is also distinct in that it applies to the independent and dependent vowels E/O as well which behaviour does not exist for the Northern virama. Its positioning w.r.t. its base is also variable unlike that of the Northern virama. Finally, the Northern virama causes ligatures whereas this is not true of the Tamil pulli.

Further, and as a result of the difference in ligature-causation mentioned above, unifying the Tamil Brahmi pulli with the Northern Brahmi virama causes problems in text segmentation. This is explained:

It is well known that Tamil orthography in general does not employ ligatures or conjoining form due to the linguistics of the language large consonant clusters. (The few exceptions are K·SSA etc in the modern Tamil script.) Therefore when implementations define the grapheme cluster / orthographic syllable for purposes of cursor placement, text selection etc, they should recognize each occurrence of CONSONANT + PULLI (i.e. Tamil Virama) as a distinct cluster on its own. (See also https://bugs.freedesktop.org/show_bug.cgi?id=40292 which I reported against LibreOffice.) This would allow cursor placement before and after this syllable and selection of this syllable alone, as in *ma|ṅ|ka|la|m* :

ம‌|‌ங்‌|‌க‌|‌ல‌|‌ம்

Likewise, Tamil Brahmi text should also have CONSONANT + PULLI treated as a distinct cluster on its own:

𑀫|𑀗்|𑀓|𑀮|𑀫்

In other Indic scripts repetitions of CONSONANT + VIRAMA as in the pattern (CONSONANT + VIRAMA +)* CONSONANT would in total be parsed as a single syllable, as in *ma|ṅga|la|m* :

4

<div align="center">म|ङ्‍ग|ल|म्</div>

Likewise in Northern Brahmi where there is no pulli but stacks/ligatures occur:

<div align="center">ब|क्‍ष|ज|ब</div>

In the current situation, as there is no distinct pulli virama character for Tamil Brahmi, applications cannot know how to correctly split grapheme clusters for cursor placement and text selection. On the other hand, with a distinct pulli virama character, the sequence CONSONANT + PULLI can distinctly be defined to represent a full grapheme cluster, whereas the regular virama of the Northern Brahmi which produces ligatures/stacks will take on the usual behaviour of (CONSONANT + VIRAMA +)* CONSONANT being defined as a single cluster.

The differences of the Tamil Pulli vs the Northern virama in Brahmi are summarized thus:

	Tamil Pulli	**Northern Virama**
Visual form	Dot	Horizontal line
Position on base	Left, right, inside	Top
Applies to	Consonants, E/O	Consonants only
Produces ligatures	No	Yes
Closes grapheme cluster	Yes	No
Usage context	Only in Tamil Brahmi	Only in Northern Brahmi

<div align="center">*§3.2. Disunification from the anusvara*</div>

While the proposed Brahmi Pulli character has the same general dot-like shape as the Brahmi Anusvara, its placement varies significantly, as it may be placed to the right, left, top-left or inside of its base as can been in the previous samples. The anusvara, in contrast, is regularly placed to the top right, as can be seen in Aśoka's third edict (ref 4, p 16):
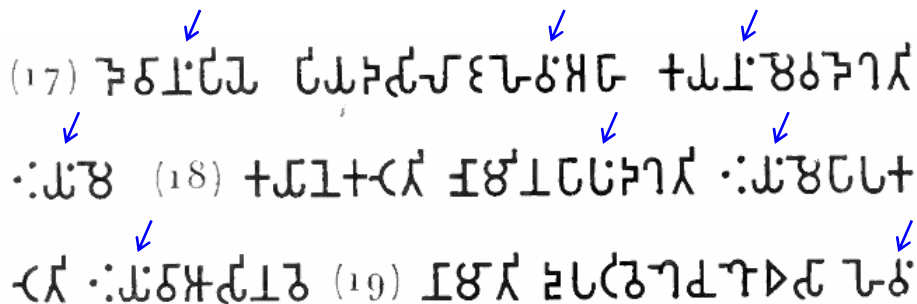
(17) devānaṃ piye piyadasi lāja hevaṃ ahā kayanaṃ meva dekhati

iyaṃ me (18) kayāne kaṭeti nomina pāpaṃ dakhati iyaṃ me pāpa ka

ṭeti iyaṃ vā āsinavē (19) nāmāti dupaṭivēkhe cu kho esā hevaṃ

cu kho esa dekhiye imāni (20) āsinavagāmīni nāma aṭhacaṃ

Technically also, its properties would be different as the proposed charactrer is a virama character with GC=9 and Indic_Syllabic_Category=Virama against the GC=0 and Indic_Syllabic_Category=Bindu of the anusvara.

Therefore the pulli is justified to be distinctly encoded for Brahmi.

§4. Usage model for Tamil Brahmi Pulli

The following model is advocated for the Tamil Brahmi pulli character:

As a vowel canceler:

╪	+	⊡		=	╪
11013 LETTER KA		11070 PULLI			

For the short vowels E/O:

▷	+	⊡		=	▷
1100F LETTER E		11070 PULLI			

⌐	+	⊡		=	⌐
11011 LETTER O		11070 PULLI			

╪	+	⊙	+	⊡		=	╪
11013 LETTER KA		11042 VS-E		11070 PULLI			

╪	+	⊙	+	⊡		=	╪
11013 LETTER KA		11044 VS-O		11070 PULLI			

Once a Tamil Brahmi pulli character is separately encoded, the existing virama 11046 BRAHMI VIRAMA would only be used outside Tamil Brahmi where a horizontal-line-shaped virama placed above is seen or ligature formation is required.

§5. Discussion of Tamil Brahmi LLA

The Brahmi proposal L2/07-342 states on p 8:

> *For the representation of sounds particular to Dravidian, the makers of Old Tamil Brāhmī added four new consonant signs to the repertoire of Brāhmī: ௴ ḻ, ௬ ḷ, ௷ ṟ and ௸ ṉ. The second of these, ௬, is phonetically identical (a retroflex lateral) to the ḷ that somewhat later appears in north-Indian Brāhmī for the writing of Sanskrit, and that also occurs in the Bhattiprolu inscriptions. Moreover, both the Tamil Brāhmī and the Bhattiprolu ḷ are graphically derived from the regular letter l, the former by adding a hook to the lower right of l, the latter by mirroring l horizontally (while the north-Indian is derived from the letter ḍ). Old Tamil, Bhattiprolu and north-Indian ḷ should therefore all be encoded as 11031.*

To be especially noted are the points where:

1) The ௬ LLA was added to Tamil Brahmi by its 'makers', just like the pulli;

2) The Northern Brahmi LLA appears only later;

3) The Tamil Brahmi LLA is derived from Brahmi LLA whereas the Northern Brahmi LLA is derived from DDA. (See also L2/12-106 §5.)

That Tamil Brahmi LLA was a separate invention from Northern Brahmi LLA in terms of originality, chronology and the glyphic base that it is derived from (i.e. LA vs DDA) should clearly give it a distinct identity from the Northern LLA. **Unicode encodes written forms and not the sounds thereof**. Thus it does not seem appropriate that the Brahmi proposal uses the term "therefore" in asserting that all three differently originated symbols that denote the phoneme /ḷa/ should be unified.

The opinion has been heard (Stefan Baums, personal communication, 2012-Feb-28) that distinct origin (i.e. Tamil Brahmi LLA deriving from LA as against DDA etc) does not justify distinct character identity. The example provided is the Sutterlin style of writing the Latin script (http://en.wikipedia.org/wiki/S%C3%BCtterlin) where A for example is glyphically derived from O but that does not mean Sutterlin A is a distinct character from regular A.

However the Sutterlin written forms were clearly intended as a new way of writing the *existing* Latin script and its letter A among others. The designer Sutterlin was certainly

entirely aware of the existence and identity of the Latin A and only devised a new glyph for it which derived from the glyph of O but which overall looks still looks like a form of A.

On the other hand, the creators of the various Brahmi written forms representing /ḷa/ were clearly entirely unaware of the existence or identity of any other representations of /ḷa/. As such there was a wilful creation of a new character identity when the various parties invented each written form for /ḷa/.

[This would mean that the Bhattiprolu LLA (and other Bhattiprolu innovations) would also be justified to be separately encoded. However, we do not propose any additional characters for Bhattiprolu Brahmi right now pending further study into such innovations.]

An excellent precedent exists in the modern Bengali/Assamese script, where the phoneme /ra/ is denoted by ৰ in Bengali and ৰ in Assamese, and these are distinctly encoded.

Thus due to the distinct identity of the Tamil Brahmi LLA ௷ compared to the Northern Brahmi LLA represented by the currently encoded 11034 BRAHMI LETTER LLA, it should be separately encoded.

Attestation for this character was already given, as in the Āṉaimalai inscription:

... the Kuṭumiyāmalai inscription:

... and so on.

§6. Collation

As these two proposed characters are meant for Tamil Brahmi, they should be collated as the corresponding modern Tamil characters 0BCD TAMIL SIGN VIRAMA and 0BB3 TAMIL LETTER LLA would be.

§7. Unicode Character Properties

```
11070;BRAHMI SIGN PULLI;Mn;9;NSM;;;;;N;;;;;
11071;BRAHMI LETTER OLD TAMIL LLA;Lo;0;L;;;;;N;;;;;
```

The codepoints chosen are the first two free codepoints in the last column of the existing Brahmi block.

The representative glyph for the Tamil Brahmi pulli character is chosen as a dot within a dashed square, to distinguish it from the existing Brahmi anusvara, and to indicate that it requires special treatment since its placement w.r.t. its base is not uniform.

Linebreak and other properties are as for the corresponding Tamil characters. Notably, the proposed BRAHMI SIGN PULLI, unlike the existing 11046 BRAHMI VIRAMA, should be treated by implementations as completing an orthographic syllable.

§8. References

1) Early Tamil Epigraphy from the Earliest Times to the Sixth Century AD, Iravatham Mahadevan, Harvard Oriental Series, Volume 6, 2003, ISBN: 0674012275

2) Indian Epigraphy, Richard Salomon, Oxford University Press, New York, 1998, ISBN: 0195099842

3) Tolkāppiyam Eḻuttatikāram, C Ganesaiyer, Cunnakam, Sri Lanka, 1937

4) Les Inscriptions de Piyadasi (Vol 2), Emile Senart, Imprimiere Nationale de Paris, 1886, http://www.archive.org/details/lesinscriptionsd02asokuoft

§9. Official Proposal Summary Form

(Based on N3902-F)

A. Administrative

1. Title
Proposal to add two characters for Brahmi
2. Requester's name
Shriramana Sharma
3. Requester type (Member body/Liaison/Individual contribution)
Individual Contribution
4. Submission date
2012-Jun-12
5. Requester's reference (if applicable)
6. Choose one of the following: This is a complete proposal (or) More information will be provided later
This is a complete proposal.

B. Technical – General

1. Choose one of the following:
1a. This proposal is for a new script (set of characters), Proposed name of script
No
1b. The proposal is for addition of character(s) to an existing block, Name of the existing block
Yes, Brahmi
2. Number of characters in proposal
2 (two)
3. Proposed category
Category C, major extinct
4. Is a repertoire including character names provided?
Yes
4a. If YES, are the names in accordance with the "character naming guidelines" in Annex L of P&P document?
Yes
4b. Are the character shapes attached in a legible form suitable for review?
Yes
5. Fonts related:
a. Who will provide the appropriate computerized font to the Project Editor of 10646 for publishing the standard?
Shriramana Sharma
b. Identify the party granting a license for use of the font by the editors (include address, e-mail etc.)
Vinodh Rajan, Shriramana Sharma, Udhaya Shankar; Open Font Licence
6a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?
Yes
6b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?
Yes
7. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?
Yes
8. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script.
See detailed proposal.

C. Technical – Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.
No
2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?
Yes
2b. If YES, with whom?

Dr R Nagaswamy, Veteran Epigraphist, Chennai; Dr G Vijayavenugopal, EFEO, Pondicherry; Stefan Baums, UC Berkeley; Andrew Glass, Microsoft; Vinodh Rajan, Chennai

2c. If YES, available relevant documents

None specifically. The matter was discussed in person and via email/phone.

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Yes. Those working with texts in Tamil Brahmi

4a. The context of use for the proposed characters (type of use; common or rare)

Common in the context of Tamil Brahmi epigraphs

4b. Reference

See detailed proposal.

5a. Are the proposed characters in current use by the user community?

They are attested in epigraphs and need to be encoded for use by those interested in Tamil Brahmi.

5b. If YES, where?

6a. After giving due considerations to the principles in the P&P document must the proposed characters be entirely in the BMP?

No

6b. If YES, is a rationale provided?

6c. If YES, reference

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

They can be placed together at the head of the column 1107x.

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

The pulli virama looks like the Brahmi Anusvara.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

Its function and properties would be different. See detailed proposal.

11a. Does the proposal include use of combining characters and/or use of composite sequences?

Yes

11b. If YES, is a rationale for such use provided?

Yes

11c. If YES, reference

The pulli virama is by nature a combining character.

11d. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

12a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

12b. If YES, describe in detail (include attachment if necessary)

13a. Does the proposal contain any Ideographic compatibility character(s)?

No

13b. If YES, is the equivalent corresponding unified ideographic character(s) identified?

13c. If YES, reference:

–o–o–o–