Response to L2/13-161 on naming Tamil fractions and symbols

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

2013-Aug-09

§1. Summary of L2/13-161

Following my revised proposal N4430 L2/13-047 to encode Tamil fractions and symbols, the Tamil Virtual Academy (TVA) has submitted L2/13-161 as feedback. I include the verbatim contents of that document as an appendix herewith for the convenience of the committees. To summarize, the TVA has expressed the opinion that the spelling of some of the proposed character names and the proposed/existing annotations should be different:

	Tamil		ISO 15919	Current	TVA
Character names proposed in L2/13-047:					
11FD2	�ௐ	சுவடு	*cuvaṭu*	Suvadu	Cuvadu
11FDD	ஹ	வராகன்	*varākaṉ*	varaaGan	varaaKan
11FDE	று	பாரம்	*pāram*	Baaram	Paaram
11FDF	கஇ	கஜம்	*kajam*	GeJam	KeSam
		கெஜம்/கெசம்	*kejam/kecam*		
11FE9	வூ	வகையரா	*vakaiyarā*	vaGaiyaraa	vaKaiyaraa
Annotations proposed in L2/13-047:					
0BF3	�	பிள்ளையார் சுழி	*piḷḷaiyār cuḻi*	pillaiyaar Suzhi	pillaiyaar Cuzhi
0BFD	சூ	சில்வானம்	*cilvāṉam*	Silvaanam	Cilvaanam
		சில்லரை	*cillarai*	Sillarai	Cillarai
Annotations already in Unicode 6.2:					
0BF9	ரூ	ரூபாய்	*rūpāy*	rUpAI	rUUpAAY
0BFD	பு	பற்று	*paṟṟu*	paTru	paRru

§2. Spelling changes proposed for character names

First let me discuss the spelling changes proposed for some character names of the newly proposed characters since they are the ones that cannot be changed post-encoding.

In my original proposal L2/12-231 p 61 I had clarified the syntax that I had followed for naming the Tamil symbols:

> For the symbols, when the meaning or significance of the character is translatable, the format is <SCRIPT> <MEANING> SIGN. as in TAMIL DEBIT SIGN etc. So we propose TAMIL BALANCE SIGN etc. When it is not translatable, the format is <SCRIPT> SIGN <NATIVE_NAME>. as in DEVANAGARI SIGN AVAGRAHA etc. So we propose TAMIL SIGN NEL etc. ... Further, in transliterating the native names, we use the popular conventions (such as zh for ḻ, etc) for readability by native users.

While my usage of the term "transliterate" is somewhat inaccurate here, neither a true transliteration nor transcription of Tamil words is feasible with the basic 26-letter Latin script alone. The TVA document speaks of "Tamil transcriptions [sic] standards that are currently in practice". I am not sure what standard(s) the TVA refers to, since:

a)   there is no reference provided;

b)   to my knowledge there is no such published standard;

c)   my inquiries* of Tamil scholars did not identify any such standards.

As such, my intention in naming these characters was to best approximate what I understand as the most common pronunciation of these words in formal Tamil such as heard in Doordarshan (TV channel founded by the GOI) Tamil News.

Now to consider the individual suggestions:

11FD2	ஹூ	சுவடு	*cuvaṭu*	**S**UVADU	**C**UVADU

Most people pronounce word-initial *c* as 's' and some as 'ch'. Hence the spelling would be preferably 'suvadu'. It is felt that 'cuvadu' would be a quite awkward spelling since this would read 'kuvadu' as per common English pronunciation outside a formal transliteration context. Note also that the TVA has not objected to the following instances of rendering Tamil ச i.e. *c* with an 's':

11FD8	டசு	பைசா	*paicā*	PAI**S**AA
11FDA	கூ	காசு	*kācu*	KAA**S**U

---

* https://listes.services.cnrs.fr/wws/arc/ctamil/2013-08/msg00025.html

11FDD	ஹூ	வராகன்	*varākaṉ*	VARAA**G**AN	VARAA**K**AN

Intervocalic stops other than *c* are normally pronounced voiced in Tamil. Again note that the TVA has not objected to these other instances:

11FD6	ஹ	பதக்கு	*patakku*	PA**D**AKKU
11FE8	மி	முதலிய	*mutaliya*	MU**D**ALIYA

11FDE	�று	பாரம்	*pāram*	**B**AARAM	**P**AARAM

While word-initial stops in native Tamil words are normally voiceless, this word is derived from Sanskrit *bhāra*[1] and is hence normally rendered with a voiced 'b'.

11FDF	கீட	கஜம்	*kajam*	**G**E**J**AM	**K**E**S**AM
		கெஜம்/கெசம்	*kejam/kecam*		

The TVA has suggested two changes here: 'g' to 'k' and 'j' to 's'. However, the word is almost always pronounced with the sounds of English 'g' and 'j', since it is derived from the Urdu word *gaz/gaj*[2]. For the same reason the primary form of the word is கஜம் transliterated *kajam*[3] (but pronounced 'gajam') although the alternate forms *kejam* and *kecam* are seen.

This would indicate that in fact this character should be named TAMIL SIGN GAJAM[4].

Another important reason that the spelling 'kesam' suggested by the TVA is dissatisfactory is that it would normally be understood as கேசம் *kēcam* which means hair. Only with the 'g' and 'j' sounds will one recognize the length measure used for cloth. All cloth merchants I have met pronounce the word with the 'g' and 'j' sounds.

---

[1] http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.8:1:5467.tamillex

[2] http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.7:1:994.platts

[3] since http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.3:1:5024.tamillex.2013941 kecam and http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.3:1:5222.tamillex.2084469 kejam both point to http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.2:1:6406.tamillex.2509528 kajam

[4] At the time of the proposal I had not suspected an Urdu derivation for this word despite it having the 'g' and 'j' sounds out of the normal places, since even some native Tamil words like calli are rendered jalli etc: http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.12:1:5851.tamillex.2316792. I shall submit a follow-up to N4430 L2/13-047 asking for the spelling to be changed to reflect the primary form of the word.

11FE9	வூ	வகையரா	vakaiyarā	va**G**aiyaraa	va**K**aiyaraa

Apart from the "intervocalic stops are voiced" point made previously in the case of 11FDD VARAAGAN, this word is also derived from Urdu* and is mostly (if not always) pronounced with the 'g' sound.

Thus I do not feel that the spelling changes suggested by the TVA either follow any transcription standard or even reflect the common pronunciation of the words. However, the current spelling of 11FDF TAMIL SIGN GEJAM should be changed to GAJAM as explained.

§3. Spelling changes proposed for annotations

The TVA has requested changes of spelling for these annotations proposed in L2/13-047:

		Tamil	**ISO 15919**	**Current**	**TVA**
0BF3	௳	பிள்ளையார் சுழி	*piḷḷaiyār cuḻi*	PILLAIYAAR **S**UZHI	PILLAIYAAR **C**UZHI
0BFD	௳	சில்வானம்	*cilvāṉam*	**S**ILVAANAM	**C**ILVAANAM
		சில்லரை	*cillarai*	**S**ILLARAI	**C**ILLARAI

The suggested spellings using 'c' for transcribing ச *c* in word-initial position are not preferable for the same reason as provided above for 11FD2 TAMIL SIGN SUVADU, i.e. the pronunciation is mostly that of 's'. While 'c' can occasionally give this sound as in 'cilvaanam', it is better to use the unambiguous 's' throughout for consistency.

Finally the following changes are requested for annotations existing in Unicode 6.2:

0BF9	௹	ரூபாய்	*rūpāy*	R**U**P**AI**	R**UU**P**AAY**
0BFD	யு	பற்று	*paṟṟu*	PA**T**RU	PA**R**RU

For 0BF9 I do think the spelling proposed by the TVA is "better" than the existing spelling. However, the one proposed for 0BFD is not in line with the existing common pronunciation of the word which is in fact pronounced closer to 'patru' than to 'parru'.

Thus among the annotations I feel the one for 0BF9 is the only one to be changed.

---

* see http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.11:1:3469.tamillex.1233335 for Tamil vakaiyarā coming from http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.9:1:2245.platts Urdu wagairā coming from http://dsalsrv02.uchicago.edu/cgi-bin/philologic/getobject.pl?c.9:1:2136.platts waġaira

Appendix 1: Contents of L2/13-161

From: Tamil Virtual Academy <tamilvu@yahoo.com>
To: Lisa Moore/Santa Teresa/IBM@IBMUS, "rick@unicode.org" <rick@unicode.org>,
Date: 07/24/2013 12:13 AM
Subject: Request to Use Tamil names in Tamil fractions and symbols in Unicode Standard
Date : 24.07.2013

Dear Dr. Lisa Moore,

Tamil Virtual Academy has gone through the documents submitted for encoding Tamil fractions and symbols for a future version of the Unicode Standard. We are happy with the progress made in encoding these rare symbols and fractions.

Because the Character names, once given, in Unicode Standard cannot be changed later and will be used by scholars and general public for many years in the future, Tamil Virtual Academy, Dept. of Information Technology, Government of Tamil Nadu requests the following Tamil names to be used for the characters. This is from the Tamil transcriptions standards that are currently in practice.

The changes needed in Reference document L2/13-047 are:

Character names:
(1) சுவடு – use CUVADU in 11FD2 as character name
(2) வராகன் - use VARAAKAN in 11FDD as character name
(3) பாரம் - use PAARAM in 11FDE as character name
(4) கெசம் - use KESAM in 11FDF as character name
(5) வகையறா – use VAKAIYARAA in 11FE9 as character name

In Annotations:
(6) சுழி – use CUZHI in 0BF3 annotation
(7) ரூபாய் - use RUUPAAY in 0BF9 annotation
(8) சில்வானம்/சில்லரை - use CILVAANAM/CILLARAI in 0BFD annotation
(9) பற்று - use PARRU in 0BF6 annotation

Thanking You
Yours faithfully,

Director,
Tamil Virtual Academy,
Department of Information Technology,
Government of Tamil nadu,
Chennai.

அன்புடன்,
தமிழ் இணையக் கல்விக்கழகம்,
காந்தி மண்டபம் சாலை,
அரசு தகவல் தொகுப்பு விவரம் எதிரில்
சென்னை – 25.
தொ.பே: 2220 1012 / 13
மின் முகவரி: tamilvu@yahoo.com
அரிச்சுவடி முதல் தமிழ் கற்க : www.tamilvu.org

Appendix 2: Response to an alternate suggestion re naming

While discussing the suggestions of the TVA with others interested in Indic encoding issues, it was suggested that the pseudo-transliteration system currently used in Indic Unicode character names may be followed for consistency. However, this is found highly dissatisfactory to native users since the reduplication of consonants (such as NA vs NNA vs NNNA) in this system would make whole words written this way quite unreadable or misleading (such as ANNAA below which suggests a geminated N where there is none):

In the case of characters which the TVA's document mentions:

		ISO 15919	Current	TVA	UCS-Pseudo-ASCII
Character names proposed in L2/13-047:					
11FD2	௵	*cuvaṭu*	SUVADU	CUVADU	CUVATTU
Annotations proposed in L2/13-047:					
0BF3	௳	*piḷḷaiyār cuḻi*	PILLAIYAAR SUZHI	PILLAIYAAR CUZHI	PILLLLAIYAAR CULLLI
0BFD	௷	*cilvāṉam*	SILVAANAM	CILVAANAM	CILVAANNNAM
Annotations already in Unicode 6.2:					
0BFD	௶	*paṟṟu*	PATRU	PARRU	PARRRRU

In the case of other characters:

		ISO 15919	Current	TVA	UCS-Pseudo-ASCII
11FD3	௹	*āḻākku*	AAZHAAKKU	-	AALLLAAKKU
11FD5	௕	*mūvuḻakku*	MUUVUZHAAKKU	-	MUUVULLLAKKU
11FD7	ௗ	*mukkuruṇi*	MUKKURUNI	-	MUKKURUNNI
11FD9	௺	*aṇā*	ANAA	-	ANNAA
11FDB	௅	*paṇam*	PANAM	-	PANNAM
11FE0	௸	*kuḻi*	KUZHI	-	KULLLI
11FE1	௴	*vēli*	VELI	-	VEELI
11FEB	ௗ	*piḷḷai*	PILLAI	-	PILLLLAI

–o–o–o–