

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/279056774>

Sri Lanka Experience of Development of Tamil Input/Output/Display Methods

Conference Paper · May 1997

CITATIONS

5

READS

45

1 author:



S. T. Nandasara

University of Colombo

107 PUBLICATIONS 224 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



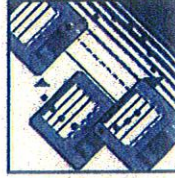
RECOGNITION OF ANCIENT SINHALA INSCRIPTION CHARACTERS [View project](#)



Language identification of web pages [View project](#)



Proceedings



TAMILNET'97

International Symposium

***for Tamil Information Processing
and Resources on the Internet***

**17-18 May 1997
Singapore**

Sri Lankan Experience of Development of Tamil Input/Display Methods

ST Nandasara, stn@sfc.wide.ad.jp, stn@ict.cmb.ac.lk

Abstract

Sri Lanka is a multi-racial society comprising of a 74% Sinhala-speaking population and a 18% Tamil-speaking population. Sinhala, Tamil and English are all official languages of Sri Lanka and are extensively spoken throughout this country of 18 million people.

In 1989, we launched our Sinhala/Tamil Word processor for DOS environment and initially it was available only in text based environment. This development of the interface for DOS was two-stage process. The first stage was to develop the BIOS part and the second was development of Text Based Word processor including Keyboard Input method and the Screen Handler.

This experience was very much helpful for the later stage to development such Input systems for the Graphical User Interface (GUI) like Windows95 and now research are being place in Sri Lanka to adapt this method. As a result Tamil Input/Display methods are available for Internet users as well.

In July 1996, Sri Lanka launched its National Website (<http://www.lk>) initially consisting of information entirely in the English language, as it is the language of commerce and government second language. Moreover, most people currently with Internet access and computer-literate are proficient in English.

However, because of our population profile, it is absolutely essential that Internet data be made available in Sinhala and Tamil in order for the Internet to reach our masses.

In addition, keyboard input technology is being developed to allow users to download, search and create new data and build their own content in all three languages. As demand increases, this will encourage information providers to leverage on the technology to put up more quality information in all three languages so that Internet technology can take root in Sri Lanka.

Introduction

The Tamil script is a South Indian script. South Indian scripts are structurally related to the North Indian scripts, but they are used to write Dravidian Languages of Southern India and of Sri Lanka, which are genetically unrelated to the North Indian Languages such as Hindi, Bengali, and Gujarati. The shapes of letters in the South Indian script are generally quite distinct from the shapes of letter in Devanagari and its related scripts. This is partly a result of the fact that the South Indian scripts were originally carved that square, block-like shapes.

The Tamil script is used to write the Tamil language of Tamil-Nadu State in India as well as minority languages such as Badaga. Tamil is also used in Sri Lanka, Singapore, and parts of Malaysia. The Tamil script has fewer consonants than the other Indian scripts. It also lacks conjunct consonant forms. Instead of conjunct consonant forms, the virama is normally fully depicted in Tamil text.

Nature of the Tamil Language

The Tamil alphabet (Tamil-alphabet ::= <Vowels><Consonants><Consonant -modifiers>) consists of 48 symbols: 22 consonants, 12 vowels and 13 consonant modifies as follows:

Table 1 A summary of the Tamil Letters

Consonants	க KA	ங NGA	ச CA	ஜ JA	ஞ NYA	ட TTA	ண NNA	த TA	ந NA	ன NNNA	ப PA	
	ம MA	ய YA	ர RA	ற RRA	ல LA	ள LLA	ழ LLLA	வ VA	ஷ SSA	ஸ SA	ஹ HA	
Vowel	அ A	ஆ AA	இ I	ஈ II	உ U	ஊ UU	எ E	ஏ EE	ஐ AI	ஒ O	ஔ OO	ஔள AU
Consonants Modifiers		ஃ AA	ஃ I	ஃ II	ஃ U	ஃ UU	ஃ E	ஃ EE	ஃ AI	ஃ O	ஃ OO	ஃள AU
	ஃ A	ஃ AA	ஃ I	ஃ II	ஃ U	ஃ UU	ஃ E	ஃ EE	ஃ AI	ஃ O	ஃ OO	ஃள AU
	ஃ VIRAMA	ஃ AU LENGTH										

Consonant modifiers in Tamil, which are graphical signs, used in conjunction with Tamil consonants. These consonant modifies can occur in left, right and top of any Tamil consonants. However no consonants modifies can occur bottom of any consonants where as Sinhala it occurs.

Special Characters

Tamil AU LENGTH MARK (ஃள) is provided as an encoding for the right side of the surroundant (or two-part) vowel sign AU (ஃள). Note that the Tamil vowel sign AU LENGTH mark (ஃள) is not the Tamil Letter LLA (ள).

Development early 1990s

The Institute of Computer Technology, University of Colombo Started development of Sinhala/Tamil Word Processor for Personal Computer under the Disk Operating System (DOS). However, this word processor system is not capable of mixing all tree language together in one document due to the limitation of DOS operating system. We have designed a bilingual font set for the display of both Tamil and English or Sinhala and English simultaneously. This was done making use of the upper extended ASCII character range for the Sinhala and Tamil characters, while retaining the basic alphabet and punctuation marks in the lower ASCII range. Language can be selected by toggling the key combination when ever is required.

In 1991, this was demonstrated in the Annual Computer Society Conference in Colombo and it allowed for the most of the e-mail users in the world to be send or to read messages in both languages simultaneously English and Tamil or English and Sinhala. System will be recognized and displayed Tamil or Sinhala font correctly when they occur, without having to change font set.

For this development, at that time we recognized 12 vowels (அ, ஆ, இ, ஈ, உ, ஊ, எ, ஏ, ஐ, ஒ, ஔ and ஔள) and 18 consonants (க, ங, ச, ஞ, ட, ண, த, ந, ன, ப, ம, ய, ர, ற, ல, ள, ழ and வ) in Tamil Language. These 30 sounds are the initial sounds of the Tamil Language and the basic for the Tamil alphabet, where as the new Tamil character set consists of more

consonants as shown in Table 1. The 18 consonants joined with 12 consonant modifiers (ா, ி, ீ, ு, ொ, ோ, ௌ and ்) form the remaining 216 glyphs for the language. The special character ஸ்ரீ (SRI) were added to the Tamil Character set, which was used in Sri Lanka for some time. Table 2 below shows the Total character set and Table 3 shows the ASCII character map for the English-Tamil font set.

Table 2 - Total Tamil Character Set

	1	2	3	4	5	6	7	8	9	10	11	12	13
Vowels	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஒ	ஓ	ஔ	ஃ
Consonant		ா	ி	ீ	ு	ூ	ெ	ே	ை	ொ	ோ	ௌ	்
Modifiers													
1	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கௌ	க்
2	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙௌ	ங்
3	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சௌ	ச்
4	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞௌ	ஞ்
5	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டௌ	ட்
6	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணௌ	ண்
7	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தௌ	த்
8	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நௌ	ந்
9	ன	னா	னி	னீ	னு	னூ	னெ	னே	னை	னொ	னோ	னௌ	ன்
10	ப	பா	பி	பீ	பு	பூ	பெ	பே	பை	பொ	போ	பௌ	ப்
11	ம	மா	மி	மீ	மு	மூ	மெ	மே	மை	மொ	மோ	மௌ	ம்
12	ய	யா	யி	யீ	யு	யூ	யெ	யே	யை	யொ	யோ	யௌ	ய்
13	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ	ரௌ	ர்
14	ற	றா	றி	றீ	று	றூ	றெ	றே	றை	றொ	றோ	றௌ	ற்
15	ல	லா	லி	லீ	லு	லூ	லெ	லே	லை	லொ	லோ	லௌ	ல்
16	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளை	ளொ	ளோ	ளௌ	ள்
17	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழௌ	ழ்
18	வ	வா	வி	வீ	வு	வூ	வெ	வே	வை	வொ	வோ	வௌ	வ்
	ஸ்ரீ												

Table 3 - ASCII character map for the English-Tamil font set

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0			SP	0	@	P	`	p			SP	ச	டி	மு	்	
1			!	1	A	Q	a	q			அ	ஞ	ட	கூ	ா	
2			"	2	B	R	b	r			ஆ	ட	கு	கு	ி	
3			#	3	C	S	c	s			இ	ண	ச	ஞூ	ீ	
4			\$	4	D	T	d	t			ஈ	த	கு	ெ	ு	
5			%	5	E	U	e	u			உ	ந	ஞ	ணூ	ு	
6			&	6	F	V	f	v			ஊ	ன	டு	தூ	,	
7			'	7	G	W	g	w			எ	ப	ணு	நூ	்	
8			(8	H	X	h	x			ஏ	ம	து	னூ	ெ	
9)	9	I	Y	i	y			ஐ	ய	நு	மு	ே	
A			*	:	J	Z	j	z			ஒ	ர	னு	ரு	ை	
B			+	;	K	[k	{			ஓ	ற	மு	றூ		
C			,	<	L	\	l				ஔ	ல	ரு	லூ		
D			-	=	M]	m	}			ஃ	ள	று	னூ		
E			.	>	N	^	n	~			க	ழ	லு	மு		
F			/	?	O	_	o				ங	வ	ஞ	று		

With this code table we make use of the over print capability (without moving cursor to the next location) and specially design consonant modifiers to keep new glyph together to generate new glyph, that is not found in the above ASCII table. For example, glyphs of row number 3 of the Table 2 can be formed with the corresponding consonant with its modifier ீ together side-by-side. Glyphs of row number 4 and 13 can be generated by over printing the consonant modifier ு and ொ. Some of the glyphs such as ே and ை were formed with the combination of ே and ை, and ௉ and ொ respectively.

Rendering of Tamil Script

The South Indic scripts function in much the same way as Devanagari, with the additional feature of two-part vowels. As in the Devanagari example, the words "TAMIL LETTER" and "TAMIL VOWEL SIGN" will be omitted where this does not cause ambiguity.

It is important to emphasise that in a font that is capable of rendering Tamil, the set of glyphs is greater than the number of Tamil Characters.

It is evident that the Tamil character set consisting of vowels, consonants and consonant modifies have clear differences, mainly with respect to the size of the characters. Some characters are much bigger than the others. Their shapes also differ. Although the basic shape of the character is curved, some parts are strait lines.

Unlike in Roman Scripts, most of the Indic language consonant modifies could be positioned at different locations around the consonants. These consonants modifies for Tamil can be classified in to three groups as follows.

<Consonant-modifiers> ::= <Left-modifiers><Right-modifiers><Upper-modifiers>

In the Tamil language, combinations of consonants and consonant modifiers produce different phonetic sounds. For example, the combination of the consonant க (KA) and consonant modifies given in Table 1 produce 12 different phonetic sound for the character க (KA). See Table 7 for these combinations.

In Tamil, it is important to emphasise that in a font that is capable of rendering combinations of Tamil script, the set of glyphs is greater than the number of Tamil Characters (See Table 4). However, the total number is fit in to 25x13-matrix and letter ஁ it is equivalent to 326 glyphs and this including the vowels, consonants and consonant modifies.

Vowel Reordering

As shown in Table 5, the following vowels are always reordered in front of the previous consonant cluster. The similar behavior is available in Sinhala as well (See Table 6).

Sinhala = ටො
Tamil = ொ ோ ௌ

Table 5 - Vowel Re-ordering in Tamil

	Key-in	Memory Representation			Display
Tamil	கெ	க	ௌ	→	கெ
	கே	க	ஔ	→	கே
	கை	க	ஐ	→	கை
Sinhala	කෙ	ක	ඌ	→	කෙ

Table 6 - Vowel Re-ordering in Sinhala

	Key-in	Memory Representation						Display
Sinhala	කෙ	ක	ඌ				→	කෙ
	කෙ ්	ක	ඌ	ඌ	ඌ		→	කෙ ්
	කෙ ් ්	ක	ඌ	ඌ	ඌ	ඌ	→	කෙ ් ්

Table 4 - Tamil Glyphs

0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	அ	ஆ	இ	ஈ	உ	ஊ	எ	ஏ	ஐ	ஓ	ஔ	ஐ	ஔ
2	ஐ	ஓ	ஔ	ஐ	ஔ	ஐ	ஔ	ஔ	ஔ	ஔ	ஔ	ஔ	ஔ
3	க	கா	கி	கீ	கு	கூ	கெ	கே	கை	கொ	கோ	கைள	க்
4	ங	ஙா	ஙி	ஙீ	ஙு	ஙூ	ஙெ	ஙே	ஙை	ஙொ	ஙோ	ஙைள	ங்
5	ச	சா	சி	சீ	சு	சூ	செ	சே	சை	சொ	சோ	சைள	ச்
6	ஜ	ஜா	ஜி	ஜீ	ஜு	ஜூ	ஜெ	ஜே	ஜை	ஜொ	ஜோ	ஜைள	ஜ்
7	ஞ	ஞா	ஞி	ஞீ	ஞு	ஞூ	ஞெ	ஞே	ஞை	ஞொ	ஞோ	ஞைள	ஞ்
8	ட	டா	டி	டீ	டு	டூ	டெ	டே	டை	டொ	டோ	டைள	ட்
9	ண	ணா	ணி	ணீ	ணு	ணூ	ணெ	ணே	ணை	ணொ	ணோ	ணைள	ண்
10	த	தா	தி	தீ	து	தூ	தெ	தே	தை	தொ	தோ	தைள	த்
11	ந	நா	நி	நீ	நு	நூ	நெ	நே	நை	நொ	நோ	நைள	ந்
12	ன	னா	னி	னீ	னு	னூ	னெ	னே	னை	னொ	னோ	னைள	ன்
13	ப	பா	பி	பீ	பு	பூ	பெ	பே	பை	பொ	போ	பைள	ப்
14	ம	மா	மி	மீ	மு	மூ	மெ	மே	மை	மொ	மோ	மைள	ம்
15	ய	யா	யி	யீ	யு	யூ	யெ	யே	யை	யொ	யோ	யைள	ய்
16	ர	ரா	ரி	ரீ	ரு	ரூ	ரெ	ரே	ரை	ரொ	ரோ	ரைள	ர்
17	ற	றா	றி	றீ	று	றூ	றெ	றே	றை	றொ	றோ	றைள	ற்
18	ல	லா	லி	லீ	லு	லூ	லெ	லே	லை	லொ	லோ	லைள	ல்
19	ள	ளா	ளி	ளீ	ளு	ளூ	ளெ	ளே	ளை	ளொ	ளோ	ளைள	ள்
20	ழ	ழா	ழி	ழீ	ழு	ழூ	ழெ	ழே	ழை	ழொ	ழோ	ழைள	ழ்
21	வ	வா	வி	வீ	வு	வூ	வெ	வே	வை	வொ	வோ	வைள	வ்
22	ஷ	ஷா	ஷி	ஷீ	ஷு	ஷூ	ஷெ	ஷே	ஷை	ஷொ	ஷோ	ஷைள	ஷ்
23	ஸ	ஸா	ஸி	ஸீ	ஸு	ஸூ	ஸெ	ஸே	ஸை	ஸொ	ஸோ	ஸைள	ஸ்
24	ஹ	ஹா	ஹி	ஹீ	ஹு	ஹூ	ஹெ	ஹே	ஹை	ஹொ	ஹோ	ஹைள	ஹ்
25	க்ஷ	க்ஷா	க்ஷி	க்ஷீ	க்ஷு	க்ஷூ	க்ஷெ	க்ஷே	க்ஷை	க்ஷொ	க்ஷோ	க்ஷைள	க்ஷ்
26	று												

Ligatures

The following examples illustrate the range of ligatures available in Tamil. These changes take place after vowel reordering and vowel splitting

1. The vowel “ஈ” optionally legates with ண, ன, or ற on its left.

ண + ஈ → ணை
ன + ஈ → னை
ற + ஈ → றை

Since this process takes place after reordering and splitting, the following ligatures may also occur:

Separate Vowels	
ண + ெ + ஈ → ெணை	
ண + ே + ஈ → ேணை	
ன + ெ + ஈ → ெனை	
ன + ே + ஈ → ேனை	
ற + ெ + ஈ → ெறை	
ற + ே + ஈ → ேறை	

2. The vowel sign “ி” and “ீ” form ligature with “ல” on their left.

ல + ி → லி
ல + ிீ → லீ

These vowels often change shapes or position slightly to link up with the appropriate shape of the consonant on their left:

ல + ி → லி
ல + ிீ → லீ

3. The vowel signs “று | று” and “று | று” typically change form of legate.

x	x + று று	x + று று
க	கு	கூ
ங	ங்	ஙூ
ச	சு	சூ
ஞ	ஞ்	ஞூ
ட	ட்	டூ
ண	ண்	ணூ

0307

த	து	தூ
ந	நு	நூ
ன	னு	நூ

x	x + ட ு	x + ு ு
ப	பு	பூ
ம	மு	மூ
ய	யு	யூ
ர	ரு	ரூ
ற	று	றூ
ல	லு	லூ
ள	ளு	ளூ
ழ	ழு	ழூ
வ	வு	வூ

4. To the right of ஜ, ஷ, ஸ, ஹ, or க்ஷ these forms have a spacing form.

ஜ + ட → ஜு
ஜ + ு → ஜூ
ஷ + ட → ஷு
ஷ + ு → ஷூ
ஸ + ட → ஸு
ஸ + ு → ஸூ
ஹ + ட → ஹு
ஹ + ு → ஹூ
க்ஷ + ட → க்ஷு
க்ஷ + ு → க்ஷூ

5. The vowel sign “ஐ” changes to “ஐ” to the left of “ண”, “ன”, “ல”, or “ள”.

ஐ + ண → ஐண
ஐ + ன → ஐன
ஐ + ல → ஐல
ஐ + ள → ஐள

Remember that this change takes place after the vowel reordering; in the first example, the vowel “ஐ” follows “ண” in the memory representation. After vowel reordering, it is on the left of “ண”, and thus changes form. The complete process is

ண + ெ → ெ + ண → ெண

Tamil Character Font Set

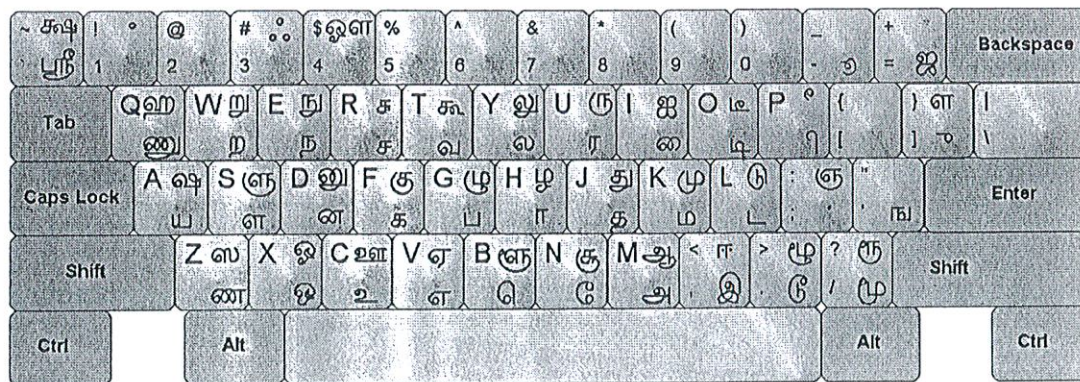
In the implementation, we have designed a font set for the display of Tamil (See Table 7). This was done by making use of the upper extended ASCII character range for the Tamil character, while retaining the basic English alphabet and punctuation in the lower ASCII range. This will allow the most of the all Tamil glyphs to displayed correctly, some using precompiled glyphs or use of the kerning feature built into the True-Type Font Technology to combined two Tamil characters into a new character glyph not found in the Font Table. With the combination of two or more Tamil character or Vowels set to give a more complex glyph, we can then include the entire Tamil character set within one single font.

Table 7 - Tamil Font Table

	8	9	A	B	C	D	E	F
0	ண்	ன்	ஷ்	ஸ்	ஹ்	க்ஷ்	ங்	ஜி
1	டி	பி	யி	லி	வி	ஷி	ஸி	க்ஷி
2	ஙீ	ஜீ	ஈ	பீ	யீ	லீ	வீ	ஷீ
3	ஸீ	க்ஷீ	ஐ	ஔ	று			
4		ஃ	ஃ	அ	ஆ	இ	ஈ	உ
5	உள	எ	ஏ	ஐ	ஒ	ஓ	ஔ	க
6	ங	ச	ஜ	ஞ	ட	ண	த	ந
7	ன	ப	ம	ய	ர	ற	ல	ள
8	ழ	வ	ஷ	ஸ	ஹ	க்ஷ	ஃ	ஃ
9	ி	ீ	ஊ	஋	ெ	ே	ை	ௌ
A	கு	ஙு	சு	ஜு	ஞு	டு	ணு	து
B	நு	னு	பு	மு	யு	ரு	று	லு
C	ளு	மு	வு	ஷு	ஸு	ஹு	க்ஷு	கூ
D	நூ	கு	ஜு	ஞூ	டு	ணூ	தூ	நூ
E	னூ	பூ	மூ	யூ	ரூ	றூ	லூ	ளூ
F	ழூ	வூ	ஷூ	ஸூ	ஹூ	க்ஷூ	ஃ	

Tamil Keyboard layout Design

Figure 1 - Tamil keyboard Layout



Conclusion

Sri Lanka National Web Site was designed to provide information in Sinhala, Tamil and English. The development of Tamil input system is equally important to the Sri Lankan society to allow for Trilingual Web Display, E-mail Interchange, Keyboard Input and other essential Internet functionalities. Trilingual Web Site allows users to view same piece of information in their preferred language, which could be English, Sinhala or Tamil. It also allows the user to search the database using their preferred language.

References

1. Govindasamy, n (1989). Keyboard for Tamil Computer, by Naa Govindasamy, 7th International Conference of Tamil Studies Seminar Proceedings, Mauritius, December 1989.
2. Sri Lanka Standard for the Sinhala Character Code for Information interchange, Sri Lanka Standard institute, Sri Lanka, 1996.
3. Report of the Committee for Standardization of Keyboard layout for Indian Script Based Computer, Electronics-Information & Planning, Vol 14, No. 1 October 1986.

Speakers'/Authors' Profile

Authors : S.T. Nandasara (stn@sfc.wide.ad.jp)

Organizations : Institute of Computer Technology, University of Colombo,
PO Box 1490, Colombo, Sri Lanka

Point of Contact: S.T.Nandasara (stn@sfc.wide.ad.jp)
Tel: 941-503150
Fax: 941-587239

URL : <http://www.ict.cmb.ac.lk>

