

Bridging the Digital Divide in Sri Lanka: Some Challenges and Opportunities in using Sinhala in ICT

S. T. Nandasara

University of Colombo School of Computing
No. 35, Reid Avenue
Colombo, Sri Lanka
stn@ucsc.cmb.ac.lk

Yoshiki Mikami

Language Observatory
Nagaoka University of Technology
Nagaoka, Niigata, 940-2188, Japan
mikami@kjs.nagaokaut.ac.jp

Abstract

The "digital divide" is the gap in technology usage and access. The digital divide has been investigated by scholars [1] and policy makers [2] mainly as an economy-specific issue that permeates the population across all demographic profiles, such as income, gender, age, education, race, and region, but not specific to the languages of different communities. The lack of native language driven ICT is a major conducive factor in digital divide.

Sinhala writing system used in Sri Lanka is a syllabic writing system derived from *Brahmi* which consist of vowels, consonants, diacritical marks and special symbols constructs. Several of these constructs are combined to form complex ligatures. The total number of different glyphs is almost close to 2300 in Sinhala language. Thus, all computer equipments that support Sinhala language needs to support a greater degree of complexity in both display and printing with near minimal changes to the keyboard or the input systems. In this paper we discuss (1) historical background of the Sinhala writing system, (2) Sinhala scripts' characteristics and complexities and illustrate (3) how Sinhala computing technology has evolved over the last quarter century. Major steps are marked by the design of character code standards as a corner stone of whole architecture for text processing. A case described in this article of "Digital Inclusion" shows how small communities of non-Roman script users can connect to the Romanized system dominated cyberspace.

1. Introduction

As far as Asia is concerned, most of the countries are far behind in developing content in their native languages. Millions of people out there who have ideas are not able to express them digitally "Digital-Contents" due to the lack of native language support whilst many are not able read digital contents due to the same reason.. In Sri Lanka; though development of multilingual computing facilities (for example Sinhala & Tamil) traced back to as early as mid 80's, a survey shows (LOP, 2007) that the web has only 0.02% of the Sinhala content pages for 1000 Sinhala speaking inhabitants..

Primarily, there are three living languages in Sri Lanka. They are Sinhala, Tamil and English, used for

general, every day communication: both interpersonal and mass communication. Of them, Sinhala and Tamil are considered "national languages" while English is considered as a "link language" to link the major ethnic groups of the Island. Thus written documents, on paper or other materials, appear in one, two or all of these languages.

Sinhala language is a syllabic alphabet in which all consonants have an inherent vowel /a/. This alphabet differs from all other Indo-Aryan languages in that it contains special sounds that are unique to it since 8th century A.D.

1. The presence of a set of five nasal sounds known as "half nasal" or "prenasalised stops" in Sinhala writing is very unique (*ṅga*, *ṅja*, *ṅḍa*, *ṅḍa* and *mba*). These five consonants have no equivalent in any Indic languages.
2. It has a pair of unique vowel symbols (*æ* and *æ̃*) to represent two vowel sounds and in use since 9th century A.D.

This article focuses on key issues and the structure concerning Sinhala writing at the character level. Then it progresses to examine the design and development of script for deferent technological generation such as printing and typewriters. This will be followed by a discussion on some of the major issues involved in design of Sinhala computing interface for early character base machines and finally philosophy behind design of 8-bit Sri Lankan ASCII (called SLASCII) and ISO 10646 code design.

2. Sinhala scripts structure and major features

Sinhala is a uniquely spoken and written native language in Sri Lanka. Sinhala script is used for writing the Sinhala language. Sinhalais said to have derivatives from the ancient scripts *Brahmi*, known to have existed since third to second century B.C.E. Subsequently alphabet and writing systems have changed considerably with notable influence by the *Kadamba* and *Pallawa Grantha* script of south India [3] [4] [5]. Full Sinhala script includes the symbols necessary for writing loan words from Sanskrit and Pali, notably the aspirated consonants.

Table 1 : Sinhala Consonant Syllabics in order of Organs and Place of Articulation [6]

	Gutturals		Palatals		Cerebrals		Dentals		Labials	
Voiceless non-Aspirate	ක	ka	ච	ca	ට	ta	ත	ta	ප	pa
Voiceless Aspirates	ඛ	kha	ඡ	cha	ඨ	tha	ච්	tha	ඵ	pha
Voiced non-Aspirates	ග	ga	ඣ	ja	ඩ	da	ද	da	බ	ba
Voiced Aspirates	ඝ	gha	ඤ	jha	ඪ	dha	ධ	dha	භ	bha
Pure-Nasals	ඞ	ṅa	ඤ, ඞ	ṅa, jṅa	ණ	na	න	na	ම	ma
Half-Nasals	භ	ṅga	ඡ	ṅja	ඬ	ṅda	ද	ṅda	ඹ	mba
Semi-Vowels			ය	va			ව	va		
Trills					ර	Ra				
Spirants			ශ	śa	ෂ	sa	ස	sa	හ, ෂ	ha, fa
Laterals					ල	la	ල	la		

There are two alphabets in current Sinhala writing system”, namely, the *E/lu*¹ alphabet and “Mixed” alphabet (see Figure 1). The *E/lu* alphabet, as described in the classical grammar *Sidatsañgarā* (ca. 1300 A.D.) comprises letters used in writing pure Sinhala words and is still widely regarded as authoritative. It contains 33 letters, of which 12 are vowels and 21 consonants. The “mixed” alphabet comprises letter of the *E/lu* alphabet and the Sanskrit alphabet. It contains 61 letters, of which 18 are vowels, 41 consonants and two diacritical marks. The mixed alphabet is use in writing *E/lu*, Pali, Sanskrit and foreign words naturalized in the language.

Table 1 (above) [6] is the Sinhala consonants syllabics laid in order of organs and place of articulation.

As illustrated in Figure 1, Sinhala differ not only in their forms and structure, but also their uses and functions i.e., in Sanskrit, Pali, classical writing, general writing, and in regional and contemporary speaking. In circumstances with Sanskrit, terms are much longer when compared with their Sinhala counterparts in modern Sinhala and also in old Sinhala. In position with writers who are committed to preserve the purity of the classical idioms, they used distinct styles, spelling and the original rules of word formation like *vid-yā-la-ya-ya* (College), which is a noun terminates with *ya-ya* is written by modernist with only single ‘ya’.

English Meaning	Old Sinhala (E/lu Alphabet)	Modern Sinhala	Classical Sinhala (Mixed Alphabet)
Household	ගහපාති (ga-ha-pa-ti)	ගෙයපාති (ge-ya-pa-ti)	ගහපාති (gr-ha-pa-ti)
Crypt	ලෙන (le-na)	ලෙන (le-na)	ලයන (la-ya-na)
Road	මග (ma-ga)	මග (ma-ga)	මාර්ග (mār-ga)
Reverent	තෙර (te-ra)	තෙර (te-ra)	ස්ථවිර (stha-vī-ra)
Teacher	අචරිය (ava-ri-ya)	ගුරු (gu-ru)	ආචාර්ය (ā-chār-ya)
Wife	බරිය (ba-ri-ya)	බිරිඳ (bi-ri-ṅda)	භාර්යා (bhār-ya-yā)
Brigadier	සෙනපිති (se-na-pi-ti)	සේනාපති (sē-nā-pa-ti)	සෙනාපති (se-nā-pa-ti)
First	පරමක (pa-ra-mu-ka)	පළමු (pa-la-mu)	ප්‍රථම (pra-mu-kha)
Minister	අමාති (a-ma-ti)	අමාති (æ-ma-ti)	අමාත්‍ය (a-māt-ya)
Buddhist Monk	සගයෙ (sa-ga-ye)	සග (sa-ga)	සංඝයා (sañ-gha-yā)

Figure 1 : Use of Sinhala in Old, Modern and Classical Society

2.1 Syllable structure

Syllables in spoken Sinhala are two types Open Syllables and Close Syllables [6]. Open syllables are composed of a vowel (V) or which ends in a vowel (CV). In V structure vowel can occur only at the beginning of a word and they can be short as *i* in *i-ra* (sun) or long as *ī* in *ī-ye* (yesterday) and single as *a* in *a-da* (today) or clusters (VV) as *iu* in *iu-va* (cooked).

CV structure occurs initially, medially and finally in words. All consonants occur in such syllables except the half-nasals and nasals (i.e. ago). For a single vowels, *ka-ta-ka-tā* (rumor), *pi-ri-si-du* (clean), *da-ha-ya* (ten) are examples. CV syllable structure can also be clusters (CVV) i.e. *ræu-la* (beard), *kæu-ma* (Sweet Bread), *du-vai* (run), *ka-rai-da* (will do it?).

A syllable which ends with a consonant is called Close syllable. It may be preceded by a vowel only (VC) or a consonant and a vowel (CVC).

¹ The term *E/lu* is given to the pure dialect of Sinhala unmixed with foreign words, and *Sirihala* to the mixed dialect, though in point of signification the two terms have not the least difference. *Sihala* in Pali, *Sirihala* in Sanskrit and *heḷa* in E/lu.

of the Sinhala letter. Some would create a rather uneven, irregular and illogical outer appearance.

Every combination is constructed in the way according to the shape of the Sinhala letter. Forty one (41) consonants (C) and sixteen (16) vowel signs (V) combined to form a glyphs. Thereafter, each united glyphs can be further combined with 2 special symbols, *rakaransaya* and *yansaya* (Cry) and then even further it can be combined with 2 diacritical marks (D) and after all it will produce more than 2300 “usable” combinations used for Sinhala writing. For example consonant *ka* (ක) with vowel signs and special symbols will produce following combinations;

ක, ක්, ක, කා, කැ, කෑ, කී, කී, කු, කු, කා,
කා, කෙ, කේ, කෙ, කො, කෝ, කො, කු,
කු, කු, කු, කී, කී, කු, කු, කො, කො,
කෝ, කො, කා, කා, කා, කා, කා, කා,
කා, කා, කො, කො, කො, කො, කො
& කො.

However, these combinations are more complicated when single or multiple vowel signs are attached to the same character. Keep the major issues outlined in this section in view related to graphical representation of character composition and combinations in Sinhala writing, it would be interesting to see how this language was developed over a fast two and half centuries until computer machine came in to operation.

3. Historical development of Sinhala writing system

3.1 Background

The oldest writing of Sinhala can be traced back to about 3rd century B.C. These are inscriptions mainly marked by either cave or rock² found in almost all part of the Island. Usually these cave inscriptions are found below the drip-ledge where the script too is protected from water. In some cases, the writing continues as one line for about forty to fifty feet from left to right and in some cases it has been written from right to left.

3.2 Evolution of Sinhala scripts

3.2.1 Era of cave and rock inscriptions

Since the Sinhala writing system is derived from the ancient North Indian scripts, Brahmi, thirty-six have been found to be appeared in Sri Lankan inscriptions from the 3rd to 1st centuries B.C. Earliest inscriptions were geometrical in shape (Figure 3). But as the time

² The University of Cambridge, England has 274 volumes of 'Epigraphica Zeylanica' with over 3000 inscriptions from Sri Lanka (that is more inscriptions than the whole of mainland China has), including one dating back to 6th century B.C. Over 2000 of these have been deciphered, indicating the consistent development of the Sinhalese language.

passed the geometric straight line scripts gradually became rounded at the edges by 1st A.D (Figure 4).

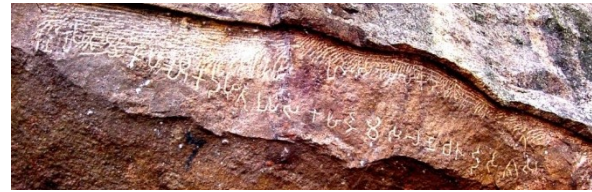


Figure 3. Vessagiri Cave Inscription in second B.C. (Source: Author’s collection, 1998)

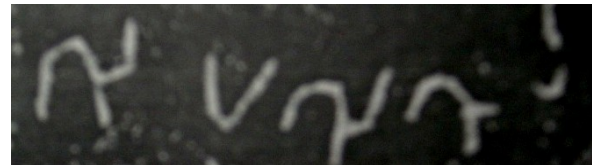


Figure 4: Cave inscription first century A.D.

3.2.2 Golden Era of Ola manuscripts

The Sri Lanka Museum in Colombo has a collection of about 3600 ola leaf manuscripts. The oldest palm leaf manuscript in existence are the *Dhampiyā Aṭuvā geṭapadaya* (belongs to the 10th century), *Chūla Vagga* (12th century), *Amāvatura* (12th century), *Saddharmaratnāvaliya* and *Pujāvaliya* (13th century). From the 13th century A.D. onwards, the production of literature becomes more prolific.

By about the 17th century a vibrant *ola* manuscripts industry was in operation. These manuscripts were written on varied subjects such as Buddhism, history, grammar, poetry, art, medicine, astrology and rituals. The efforts of some of these highly talented scholars have been referred to both in the historical and epigraphical records of the Island.

Sinhala characters’ rounded shape as highlighted earlier and the present shapes of the characters evolve mainly due to the use of *ola leaf* (from *Palmyra* tree) from the very early time. Use of *ola* leaf with sharper steel stylus point, gradually took the present rounded shape to form the modern Sinhala alphabet (Figure 5).

However, Portuguese rule came to an end in 1658, and next European power, Dutch became the masters of the coastal districts (1656-1796). Dutch maintained those records (tombōs) and in addition they also made a more important contribution of charting the area on maps. Dutch started school for Europeans and also for local people. In these schools, language of medium was on their own mother tongue. Seminaries were established in *Jaffna* (City of the northern parts of Sri Lanka) in 1760s for higher education, where Sinhala and Tamil were special subjects and hence, these educational activities demanded the need of books in their own languages. This caused the necessity of printing press to be established in Sri Lanka.

		3-1 B.C.	1-3 A.D.	4-5 A.D.	6-7 A.D.	8-10 A.D.	12 A.D.	15 A.D.	1737	1876	1891	1996 SLSI	1998 UNICODE	Notes
◌	m								◌	◌	◌	◌	◌	3-1 B.C. (Inscriptions)
◌	h										◌	◌	◌	¹ Periyankulama (207-197 B.C.)
අ	a	අ ¹	ආ ³	ඈ ³	ඉ ⁴	ඊ ⁵	උ	ඌ	අ	අ	අ	අ	අ	² Mihintale (207-197 B.C.)
ආ	ā	ආ	ඈ			ආ ⁴			ආ	ආ	ආ	ආ	ආ	³ Situlpawwa (161-137 B.C.)
ඇ	æ					ඇ ²			ඇ	ඇ	ඇ	ඇ	ඇ	⁴ Korawakgaka (77-63 B.C.)
ඈ	æ					ඈ ²			ඈ	ඈ	ඈ	ඈ	ඈ	⁵ Ritigala Wewelanne (22-7 B.C.)
ඉ	i	ඉ ¹	ඊ ⁴	උ ¹	ඌ ²	ඍ ⁵	ඎ	ඏ	ඉ	ඉ	ඉ	ඉ	ඉ	⁶ Yatahalena Vihara (22-7 B.C.)
ඊ	ī						ඊ		ඊ	ඊ	ඊ	ඊ	ඊ	⁷ Gallena Vihara (22-7 B.C.)
උ	u	උ ²	ඌ ⁵	ඍ ³	ඎ ⁴	ඏ ⁶	ඐ	එ	උ	උ	උ	උ	උ	⁸ Nuwaragala (22-7 B.C.)
ඌ	ū		ඌ ¹						ඌ	ඌ	ඌ	ඌ	ඌ	⁹ Ritigala Andiyakanna (22-7 B.C.)
ඍ	r									ඍ	ඍ	ඍ	ඍ	¹⁰ Boowattegala (22-7 B.C.)
ඎ	ṛ									ඎ	ඎ	ඎ	ඎ	¹¹ Rajagala (44-22 B.C.)
ඏ	l									ඏ		ඏ	ඏ	1-3 A.D. (Inscriptions)
ඐ	l̥									ඐ		ඐ	ඐ	¹ Anuradhapura (1-7 A.D.)
එ	e	එ ⁵	ඊ ⁶	උ ³	ඌ ³	ඍ ⁵	ඎ	ඏ	එ	එ	එ	එ	එ	² Situlpawwa (1-7 A.D.)
ඊ	ē								ඊ	ඊ	ඊ	ඊ	ඊ	³ Maharatmale (7-18 A.D.)
උඌ	ai								උඌ	උඌ	උඌ	උඌ	උඌ	⁴ Wallipuram (67-111 A.D.)
ඍ	o	ඍ ⁶			ඍ ²	ඎ ⁵	ඏ	ඐ	ඍ	ඍ	ඍ	ඍ	ඍ	⁵ Viharagala (60-67 A.D.)
ඎ	ō								ඎ	ඎ	ඎ	ඎ	ඎ	⁶ Pahala Kainattama (60-67 A.D.)
ඏඐ	au									ඏඐ	ඏඐ	ඏඐ	ඏඐ	4-5 A.D. (Inscriptions)
ක	ka	ක ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ක	ක	ක	ක	ක	¹ Tonilaga (301-328 A.D.)
ඛ	kha	ඛ ⁴			ඛ ⁴	ඌ ⁶	ඎ	ඏ	ඛ	ඛ	ඛ	ඛ	ඛ	² Ruwanweliseya (337-365 A.D.)
ග	ga	ග ²	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ග	ග	ග	ග	ග	³ Tissamaharama (406-428 A.D.)
ඝ	gha	ඝ ⁷	උ ²			ඍ ¹	ඎ	ඏ	ඝ	ඝ	ඝ	ඝ	ඝ	⁴ Anuradhapura (437-452 A.D.)
ඞ	ña						ඞ		ඞ	ඞ	ඞ	ඞ	ඞ	6-7 A.D. (Inscriptions & Pillars)
ඟ	ṅga							ඟ	ඟ	ඟ	ඟ	ඟ	ඟ	¹ Kandanadu (517-518 A.D.)
ච	ca	ච ¹	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ච	ච	ච	ච	ච	² Dhakshinathupa (639-650 A.D.)
ඡ	cha	ඡ ⁶				ඡ ⁶	ඎ	ඏ	ඡ	ඡ	ඡ	ඡ	ඡ	³ Baron Paviliyan (639-650 A.D.)
ඣ	ja	ඣ ⁸	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ඣ	ඣ	ඣ	ඣ	ඣ	⁴ Kuchchaweli (639-650 A.D.)
ඤ	jha	ඤ ²					ඞ	ඏ	ඤ	ඤ	ඤ	ඤ	ඤ	⁵ Murutawa (639-650 A.D.)
ඥ	ña						ඞ	ඏ	ඥ	ඥ	ඥ	ඥ	ඥ	8-10 A.D. (Inscriptions & Pillars)
ට	jña							ට	ට	ට	ට	ට	ට	¹ Thiriyaya (733-771 A.D.)
ඨ	ñja								ඨ	ඨ	ඨ	ඨ	ඨ	² Viyaulpotha (853-887 A.D.)
ච	ta	ච ²	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ච	ච	ච	ච	ච	³ Dorabewila (915-923 A.D.)
ඡ	tha	ඡ ⁹			ඡ ¹		ඎ	ඏ	ඡ	ඡ	ඡ	ඡ	ඡ	⁴ Baddulla (946-954 A.D.)
ජ	da	ජ ¹¹	උ ³	උ ³	ඌ ³	ඍ ⁴	ඎ	ඏ	ජ	ජ	ජ	ජ	ජ	⁵ Polonnaruwa (982-1029 A.D.)
ඣ	dha	ඣ ³	උ ³	උ ³	ඌ ³	ඍ ³	ඎ	ඏ	ඣ	ඣ	ඣ	ඣ	ඣ	⁶ Indikatuseya (982-1029 A.D.)
ඤ	na	ඤ ²	උ ³	උ ³	ඌ ⁴	ඍ ⁴	ඎ	ඏ	ඤ	ඤ	ඤ	ඤ	ඤ	12 A.D. (Inscriptions & Pillars)
ඞ	ñda									ඞ	ඞ	ඞ	ඞ	15 A.D. (Inscriptions & Pillars)
ත	ta	ත ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ත	ත	ත	ත	ත	1737 (Printed Characters)
ථ	tha	ථ ¹⁰			ථ ¹	ඌ ⁴	ඍ ⁶	ඎ	ථ	ථ	ථ	ථ	ථ	<i>Sinhala in First Printed Book (1737)</i> (see Figures 5 for sample page)
ද	da	ද ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ද	ද	ද	ද	ද	1876 "ALFABETE DES GESAMMTENERDKREISES"
ධ	dha	ධ ²	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ධ	ධ	ධ	ධ	ධ	<i>("Alphabet of all race of the world")</i> which publishes the letter printing type of the world (K.K. HOF- UND STAATSDRUCKEREI IN WIEN, 1876) by The Royal Print Shop in 1876, Vienna, Germany (29 sheets).
ඞ	ñda	ඞ ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ඞ	ඞ	ඞ	ඞ	ඞ	1891
භ	pa	භ ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	භ	භ	භ	භ	භ	<i>Alphabet listed in "A Comprehensive Grammar of the Sinhala Language" by A. M. Gunasekara. (1891)</i>
ඵ	pha						ඎ	ඏ	ඵ	ඵ	ඵ	ඵ	ඵ	
භ	ba	භ ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	භ	භ	භ	භ	භ	
ඞ	bha	ඞ ⁷	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ඞ	ඞ	ඞ	ඞ	ඞ	
ම	ma	ම ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ම	ම	ම	ම	ම	
ඞ	ba								ඞ	ඞ	ඞ	ඞ	ඞ	
ය	ya	ය ¹	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ය	ය	ය	ය	ය	
ර	ra	ර ²	උ ³	උ ³	ඌ ⁴	ඍ ⁵	ඎ	ඏ	ර	ර	ර	ර	ර	
ල	la	ල ²	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ල	ල	ල	ල	ල	SLSI Character Set 'Sarasavi' Font – 10 points (1996), Characters in Gray cells were not included due to they can be produced by combining with consonant modifiers
ව	va	ව ²	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ව	ව	ව	ව	ව	
ශ	śa	ශ ¹			ශ ⁴	ඍ ⁶	ඎ	ඏ	ශ	ශ	ශ	ශ	ශ	
ස	sa	ස ²	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ස	ස	ස	ස	ස	
හ	ha	හ ¹	උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	හ	හ	හ	හ	හ	UNICODE Character Set 'Iskoola Pota' Font – 10 points (1998)
ල	la		උ ³	උ ³	ඌ ⁴	ඍ ⁶	ඎ	ඏ	ල	ල	ල	ල	ල	
ඞ	fa								ඞ	ඞ	ඞ	ඞ	ඞ	

Figure 5: Illustration of the Evolution of Sinhala Script

The establishment of a printing press in Sri Lanka was seriously taken up first by Governor *Jacob Christian Pielat* in 1734. As a result, first Sinhala printing office was established in the city of Galle, southern part of Sri Lanka during the period of Governor *Van Imhoff* (1736-1740).

The first book published in 1737 by *Gabriel Schade* was the Sinhala prayer book (41 pages), the first of any size ever printed in Sinhala (see Figure 6). Starting of the printing trade in Sri Lanka by the Dutch was an important step taken towards the literature production in the country.

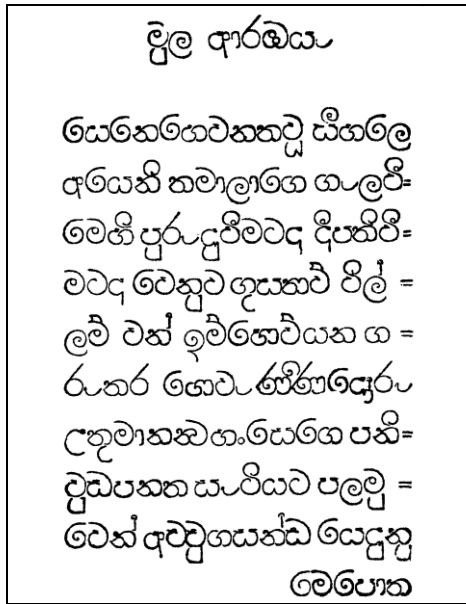


Figure 6: One Page from the First Book Printed in Sri Lanka (1737) (Source: Department of National Archives, Government of Sri Lanka, Original Image was scaled down to 46%)

Likewise, in Bengal, when the Serampore³ Missionaries founded by William Carey in 1800 (Carey's, 1993), he realised the complexity of socio-cultural problems due to its multilingual characters. William Carey and his Serampore Missionaries engaged in the development and improvements of printing industry for all the popular Indian languages including Sinhalese. As a result Sinhala printed letter was improved by its circularity and individuality. The early version of specimens of the versions of the sacred scriptures are of Asian languages including CINGALESE (Sinhala), TAMUL (Tamil), PERSIAN (Persian) etc. printed by Serampore Missionaries in 1813 (Figure 7).

In 1876, the comprehensive samples register entitled "ALFABETE DES GESAMMTEN ERDKREISES" ("Alphabet of all race of the world") which was published the letter printing type of the world [7] by The Royal Print Shop in 1876, Vienna, Germany (29 sheets). The printing type of 76 languages recorded in this sample register, among those Sinhala has been recorded as CINGALESISCH

(Figure 8) shows the European interest in Sinhala scripts. It is also important to note that the font width was very significant factor for printing even at that time, and therefore due attention was paid on the width of the each font. Breakdown of the printing type which is published in these documents were found as 13 vowels, 34 consonants, 1 semi consonant, 13 vowel signs and 41 ligatures. The 5 vowels 5 consonants, 2 consonants modifiers, 1 semi consonant and some important ligatures were not found in this list though they are vital for printing.

3.2.3 Printing establishments in Sri Lanka

Up to the time of *Colebrooke Commission* (1930), there were few printing establishments. These include the Government Printing Establishment and a few other presses run by English Missionary Societies. The growth of newspapers published in the island was a result of a historical evolution when *William Colebrooke*, who was appointed by the British Government to look into the affairs of administration of Sri Lanka on 11th April, 1829 and he himself recommended the necessity of commencing newspapers. As a consequence of the recommendation made by the Colebrooke Commission, Governor *Wilmot Horton* who arrived in the Island on 23rd October, 1831 established the newspaper Ordinance No.5 of 1839 and was notified to commence newspapers into operation.

The first Sinhala newspaper in Sri Lanka commenced publication in the country was "Lanka Loka" at the city of Galle in June 1860. This newspaper was not registered under the ordinance by the publisher *W.E. Eaton*, therefore, the first registered newspaper under the ordinance was "Lakmini Pahana" which commenced publication on 17th September, 1862. It is seen that the whole newspaper was printed using only one type face and one size of type. Even headlines were set using the font used for the text. "Gnānārta Pradeepaya" Catholic newspaper was published in 7th July, 1866 used larger font size for headlines. The Sinhala Printing was well established in Sri Lanka by the mid 19th century.

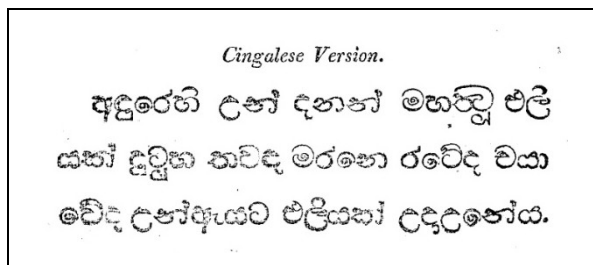


Figure 7: Specimens of Translations of the Version of the Sacred Scriptures (1813)

³ Serampore was then a Danish colony and it was small town known also as Fredricksnagore after the name of King Fredrick V of Denmark.

4. Major steps in Sinhala text processing

4.1 The first Sinhala on computer screen

With the introduction of BBC microcomputers to the University of Colombo in 1982, a set of Sinhala Bitmap fonts for computers (Figure 10) was developed. Using this Sinhala font set, daily TV programme schedule was transmitted for public by the National TV Station (Independent Television Network-ITN) and it was the first attempt to use computers with local languages. In November, 1982, the same facility was used to transmit general election result on SLRC (Sri Lanka *Rupavahini* Corporation), which is the main national television station. This was the major breakthrough and created awareness about computers are capable of using languages other than English in Sri Lanka.

Vowels	අ ඉ ඊ උ උ ශ ඊ ඊ ඊ
Consonants	ක නි ග ස ඩ භ ච ඡ ජ ඣ ඤ ඞ ට ඨ ධ ජ ඣ ඤ ඞ ට ඨ ධ ජ ඣ ඤ ඞ ට ඨ ධ ජ ඣ ඤ ඞ ට ඨ ධ ජ ඣ ඤ ඞ ට ඨ
Consonant Modifiers	ආ භ භ භ භ භ භ භ භ භ භ භ භ භ භ
Ligatures	ඡ ඡ
Consonants modifier shapes	ආ භ භ භ භ භ භ භ භ භ භ භ භ

Figure 10: First set of Sinhala characters were created for computers (SinhalaInet Font © S. T. Nandasara, 1989)

4.2 Early unsuccessful attempts

Introduction of IBM PCs for data processing and the need of developing proper applications were major challenges in computer-processing languages like Sinhala. Existing technologies lagged far behind to handle such complex scripts. The very first Sinhala word processor, developed by a Chinese company in 1984, was not successful in Sri Lanka because of the input method's unacceptable behavior (such as having to use two keystrokes for single characters like *na* [න], and the 12 line, 40 column character display). Thereafter, another Sinhala word processor, developed by the GIST (Graphics and Intelligence-based Script Technology) group in India, needed an additional hardware pluh-in board to display 80 columns and 24 lines of Sinhala text. Further, the GIST input method was based on the phonetic keyboard, which was not acceptable to Sri Lankan. This effort, too, was not successful.

Some local computer vendors who were interested in developing software for the IBM compatible personal computer ended up in a patent disputes over the software that one company developed. This consequently led to a major setback for the development of Sinhala language computing in Sri Lanka. Ultimately, this challenge was settled out of court after more than two years of litigation.

4.3 Long debates on alphabet and alphabetical order

Since mid 1980s, a number of steps were taken by the government to formulate Sinhala language related discrepancies. One of such discrepancies was different alphabetical orders were used by different dictionaries [9]. Working Committee was formed on the Use of Sinhala and Tamil in Computer Technology by the Information and Technology Council of Sri Lanka (CINTEC) in 1985. One of the committee's initial endeavors was to establish a standard code for information interchange in Sinhala and work progressed on the alphabetical order and standard codes, the associated problems of character addition and the essential feature and shape of each character.

These works were presented to the 9th Annual Conference of Computer Society of Sri Lanka (CSSL) [10] [11]. A new character to denote *fa* was introduced formally to the standard as a last character in the set.

4.4 Era of SLASCII – Typewriter metaphor

During the late-1980's, the first author made several visits to Thailand to study that countries' language development for information processing⁵. At this stage the aim was to develop an 8-bit code to fill the positions from A0 to FF in the single byte ISO 8859 like code table based on the keyboard's character set.

As a result of the collaborative work with the Thammasat University and the inputs from the CINTEC Working Committee the draft standard was released as a CINTEC publication [12] to the general public for comments and observations in March 1990.

After receiving the public comments and recommendations, the first ever encoding was approved and published by the Council of CINTEC. The 8-bit code table defining 65 Sinhala characters in the A1-CF, E0-EF and F4-F6 [13] was submitted to the Sri Lanka Standards Institute (SLSI) as the Sri Lanka Sinhala Standard Code for Information Interchange (SLASCII) for implementation.

4.5 Standard keyboard for Sinhala

At this stage, it is important to indicate that for the development of the appropriate electronic keyboard layout where again CINTEC took the initiative. Having agreed that a large number of Sinhala typists were using the government approved *Wijesekera*⁶ Sinhala Typewriter Keyboard (Figure 11), CINTEC first developed and obtained government approval for the "Extended *Wijesekera* Keyboard for Electronic Typewriters" (see Figure 12), the intention being the introduction of electronic typewriters then used as an

⁵ In this study, S. T. Nandasara closely work with Dr. Thaweesak Konantakol, Information Processing Institute for Education and Development (IPIED), Thammasat University, Bangkok, Thailand.

⁶ *Wijesekera* Typewriter Keyboard was approved by the government of Sri Lanka as a National Sinhala Typewrite in 1968.

Later, IBM also showed the Unicode code table which was published in the first publication, Unicode 1.0 in 1991.

Sinhalese ISO 10646
Draft Int. Standard

Table 40

G = 032 P = 032 R = 165

Cell	160	176	192	208	224	240
+0	ඉ	ඊ	උ	ඌ	ඍ	ඎ
+1	ඏ	ඐ	එ	ඒ	ඓ	ඔ
+2	ඔ	ඕ	ඖ	඗	඘	඙
+3	ක	ඛ	ඛ	ඞ	ඟ	ච
+4	ඛ	ඞ	ඞ	ඟ	ච	ඡ
+5	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+6	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+7	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+8	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+9	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+10	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+11	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+12	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+13	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+14	ඣ	ඤ	ඦ	ට	ඨ	ඩ
+15	ඣ	ඤ	ඦ	ට	ඨ	ඩ

-112-

Figure 14: Proposed Sinhala Code Page for ISO/IEC 10646 as expert contribution from Ireland (1989, 1996)

Formulation of this Sinhala Unicode Standard was based upon the proposal submitted by the early contributor from Ireland [19] [20]. This code page represented a distorted Sinhala character set with several glaring errors and omissions and it was based on an early grammar book [21].

Immediate steps were taken to request ISO/IEC/JTC1/SC2 and UTC directly through the Sri Lanka Standards Institute (SLSI) to suspend approval of the draft until representations were made by the CINTEC and SLSI. The work in Sri Lanka regarding the standard code was thereafter speeded up.

Meanwhile a Unicode based Sinhala Standard was formulated by the CINTEC and thereafter by a SLSI Committee [13]. The SLS 1134:1996 standard had been prepared to fall in line with the requirements laid down in ISO/IEC 10646 and maintains the logical sequence of the alphabet. Finally the Sri Lanka Standard Sinhala Character Code for Information Interchange [22] was approved by the Sectoral Committee on Information Technology and was authorized for adoption and publication as a Sri Lanka Standard by the Council of the Sri Lanka Standards Institute on 1996-09-19 and sent to the ISO/IEC/JTC1/SC2 as a Sri Lankan proposal for the Unicode Standard.

This proposal was discussed at the Singapore ISO/IEC JTC 1/SC 2/WG 2 Meeting #32 on 1997-04-01 and was recommended to examine in detail with Sri Lankan national body representatives in coming meetings.

4.9 The ISO-IEC 10646 Standard

In June, 1997, CINTEC with the assistance of NARESA two national representatives⁹ were sent to the ISO/IEC JTC 1/SC 2/WG 2 Meeting #33 held in Crete, Greece where the draft Sinhala Code was discussed intensively. National delegates argued for the draft submitted by Sri Lanka opposing several competing proposals from the Unicode Inc., USA [23], Ireland [20], UK [24] and Sri Lanka [5] [22]. After few ad-hoc committee meetings with National delegates and other nominated country delegates concluded to accept the repertoire, names, and arrangement for Sinhala script [25] based on Sri Lankan proposal with slight modification with the support of the majority delegates from Canada, Greece, United Kingdom, United States of America & Japan [26] [27]. This was ratified¹⁰ at the WG 2 Meeting #34 on 16th March, 1998 held at Redmond, Seattle, USA and the Sinhala Code Chart was included in the Unicode Version 3.0 [28]. The SLS 1134:1996 was also accordingly modified.

The Standard Sinhala keyboard layout which was first standardized in 1996 was modified to some extent in 2001. In 2004 it was further modified [29] and now it is used as the national standard keyboard layout for Sinhala.

4.10 Current development platform status

The inclusion of Sinhala code page in UCS/Unicode has made it possible to connect Sinhala community to global cyberspace. But in order to be really connected, localization should be done on proprietary or open platforms accordingly. Currently Microsoft Windows platform is widely used in Sri Lanka. Microsoft does provide proper Unicode support input method for Sinhala. Sinhala language kit released by Microsoft for Windows 7 can be used with Sinhala *Wijesekara* keyboard.

4.11 Use of Internet and local language support

Although Sri Lanka only had Internet connectivity since 1995, it has already had a significant impact on the country. The first Internet e-mail service, *LEARNmail*, was inaugurated by the University of Moratuwa in April 1990. The Sri Lanka Trilingual National Web Site was inaugurated in August 1996 [17].

The government of Sri Lanka began to focus on IT issues in the mid-1980. After Unicode had been established in 1998, unlike Thailand, India and Nepal government work closely to develop local version of Microsoft OS has not the agenda item in Sri Lanka. To overcome this problem, government must have solid policy and broad active plans to invest on local language development not only in their own soil, but

⁹ S. T. Nandasara and J. B. Disanayaka were the two members attended for WG2 #33 from Sri Lanka.

¹⁰ S. T. Nandasara was attended for WG2 #34 at Seattle, USA.

also continues long term lineup work agreements with such corporation.

Various local industries in Sri Lanka are making use of e-mail and the web for their business. However, it is notable that virtually none of them have truly integrated local language into their business due to the lack of confidence of local language products available. This is a common phenomenon not only in business sector or private sector organization but also with government and corporate sector organization, more likely; their working environment is more towards English speaking only.

In addition, non-standard keyboard input technology is being developed to allow users to insert, download, search, and create new data and build their own content in Sinhala language and use of roman text to represent Sinhala words and sentences (for sending SMSs and online chat rooms) are widely spread. Since the demand increases, this may encourage local information/Internet service providers to design proper and standard technology to put up quality information in all three languages so that Internet technology can take root in Sri Lanka. It is to be noted that development of such technologies is not in ISPs general agenda.

5. The Research infrastructure

The local language research infrastructure in Sri Lanka is very much geared towards short term objectives. Staffs in Universities have to rely principally on Masters level research students to help them with their research. As a consequence, building up research staff within any long term research programme is almost impossible. The Academics felt they could do use direct Masters level students in short three month projects, and that as soon as the students were at a level where they might do something useful, they graduated and left, leaving the supervisors to start the whole process over again. This situation is compounded by Sri Lankans doing their PhDs abroad where invariably they are steered away from working on issues related to local languages. Their research field will be selected according to the western interest. All most all the Sri Lankans who did their PhDs in Western, European, Scandinavian or even Eastern countries had not worked on localization field. The National Science Foundation does have funds available for IT research, but it is a diminutive amount of money, and with a policy of funding as many projects as possible, the emphasis is on very small, very short term projects. As a consequence longer term investment in large scale projects in IT research until recently is almost unknown. Their aid and grants for longer term investment in other areas – for example the British, Swedish, Norway and Japanese Governments funds are for long term investment in environmental research, infrastructure development, human resource development, – but not for local language research related or even not for IT related research. Recreantly, JICA and SIDA start funding in research infrastructure development in the universities, but objectives are not necessary limited

to localization research and development. Thank to IDRC's localization initiatives and funding helps University of Colombo School of Computing to establish a Local Language Research Center in 2004. Private industry has banded together to create the Software Vendors Association, but this association, while having research funds available, focuses those funds on matters such as local area networks and other basic items of commercial infrastructure. Language processing is not on their list of priorities any more.

6. Digital Divide

Apart from the language divide, Sri Lanka faces many other divides revolving around ethnicity, religion, rural/urban etc. The emergence of an information society scattered and loosely connected, and created the rapid surge in the information and communication technologies. But the slow pace with which the Sri Lankan society is trying to absorb these technologies through its organs such as language has added divide to the society in which many already existing "digital divide" resulting in the disparity in access to information and communication technology in the 21st Century. The Sinhala language in which the internet search engines can search is very small percentage. So, since what people want in the digital world is not available in their languages, both the government and the people are fast moving towards introducing English at the earliest level in their education, will be resulted to list clearly than today it is another dieing language. The language liveliness - capacity of a language to live, grow, and develop - depends upon various factors. Some of these are: social status, demography, and institutional support. Access to Information and Communication Technology in their own language is one of the ways to empower the people and enhance the vitality of a language.

Though gradually internet is evolving as a mass media around the world, the entry of information technology in Sri Lanka is not mass oriented, like other media such as news papers, book publishing, radio, television etc. It is limited in orientation, and it is also government-oriented. Though much is said, less of what is actually delivered to the end user. A lot needs to be done, and which the Sinhala language is yet to be approached by the technology initiatives. However, only localization approaches may not be successful enough to bridge the digital divide. The digital divide has to be discussed not only in the context of linguistic issues, but also in the context of other technical issues, since it involves convergence of many related technologies, the economic resources (per capita) available with people to take steps to cross the digital divide and so on.

Let us remember that everything that English has, or for that matter, what Sinhala is going to have, cannot be or may not be appropriated by other languages used in Sri Lanka i.e. Tamil, Pali, Sanskrit etc., but certainly they can increase their vitality by becoming part of the IT world in as many possible

ways as they can than being left out of the race. At present most of them are out of the race.

7. Conclusion

Language identity and its recognition has to be maintained. This can be done through local radio or TV stations. It could also be done on a low cost technology [30] and word wide scale with help of UNESCO, which is the organization committed to preserve human culture and languages and narrow down the language digital divide [31]. Education is another way of promoting and preserving languages as the means of digital divide. How should we best teach and provide learning opportunities for among human languages. Building multilingual word dictionaries, maintaining common social context and common oral corpus would be useful among the multilingual communities. Thus, the creation of digitized corpora is a very basic task in the effort to preserve the world's languages. The corpora can be multimodal, spoken or written, depending on what type of linguistic material and recording equipment is available. This will bring to a discussion of the role of language technology once again. Probably a key factor here is reuse of technology for similar languages. This will bring the solution as mentioned in the publication 'E-commerce and Development Report 2003' by the United Nations Conference on Trade and Development (UNCTAD) provides an insight into the software that developing countries can use for bridging the digital divide. Report recommends use of free or open source software as against the proprietary license-to-use software.

Acknowledgement

The study was made possible by the Asian Language Resource Network Project of the Nagaoka University of Technology with financial support of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] Hoffman, Donna L.; Novak, Thomas P., 1998. *Bridging the Digital Divide: The Impact of Race on Computer Access and Internet Use*. Educational Resources Information Center, Department of Education, U.S.A.

[2] NTIA, 1999. *Falling Through the Net: Defining the Digital Divide*, National Telecommunications and Information Administration, U.S. Department of Commerce, U.S.A.

[3] Florian Coulman, *The Writing Systems of the World*, Blackwell Publishers Ltd., 1989

[4] Florian Coulman, *The Blackwell Encyclopedia of Writing Systems*, Blackwell Publishers Ltd., 1996

[5] Fernando, P. E. E., 1949. *Palaeographical Development of the Brahmi Script in Ceylon from 3rd Century B.C. to 7th Century A.D.*, University of Ceylon Review 7.

[6] Disanayaka, J. B. *The Structure of Spoken Sinhala-Sound and their Patterns*, National Institute of Education, Sri Lanka, 1991.

[7] K.K. HOF- *UND STAATSDRUCKEREI IN WIEN*, 1876

[8] Monotype, *Book of Information*. New ed. 1959, London.

[9] Working Paper, 1985. *Order of Alphabet and System of Transliteration*, CANLIT & NARESA.

[10] Samaranyake, V. K., Disanayaka, J. B. and Nandasara, S. T., 1989. *A Standard Code for Sinhala Characters*, Proceedings, 9th Annual Sessions of the Computer Society of Sri Lanka, Colombo.

[11] Nandasara, S. T., Disanayaka, J. B., Samaranyake, V. K., Seneviratne, E. K., and Koannantakool, T., 1990. – *Draft Standard for the Use of Sinhala in Computer Technology*, Approved by the CINTEC on the advice of its working committee for recommending Standards for the Use of Sinhala and Tamil Script in Computer Technology.

[12] Nandasara S.T., 1990. “*Working Group Paper, Draft Sri Lankan Standard proposal for Sinhala Character Code for Information Interchange*”, Working Group for Sinhala Code for Information Interchange, Coordinated and Managed by Computer and Information Technology Council of Sri Lanka,

[13] Samaranyake, V. K., and Nandasara, S. T., 1990. *A Standard Code for Information Interchange in Sinhalese*, ISO-IEC JTC1/SCL/WG2 N673.

[14] Nandasara, S. T., 1991. “*Proposed Sri Lankan Sinhala Standard Code for Information Interchange (SALASCII)*”, Approved for the Computer and Information Technology Council of Sri Lanka (CINTEC) and Submitted to the Sri Lanka Standards Institute.

[5] Samaranyake, V.K. Nandasara, S.T., 1994. *A Standard Code for Information Interchange in Sinhala*, Proceedings of the International Conference on the Input and Output of National Character Sets, AFSIT-8, Tokyo, Japan.

[16] Nandasara, S.T., 1997. *Sri Lanka Experience of Development of Tamil Input/Output/Display Methods*, TAMILNET'97 – International Symposium, Singapore.

[17] Nandasara, S. T., Leong, K. Y., Samaranyake, V. K., and Tan, T. W., 1997. - *Trilingual Sinhala Tamil English National Web Site of Sri Lanka*, INET97, http://www.isoc.org/inet97/proceedings/EI/E1_3.HTM.

[18] Nandasara, S. T., Samaranyake, V. K., 1997. *Current Development of Sinhala / Tamil / English Trilingual Processing in Sri Lanka*, MLIT-2, November 7-8, Tokyo, Japan.

[19] Michael Everson, 1989. *Proposal for encoding the Sinhala script in ISO/IEC 10646 (revision 1)*. <http://www.evertype.com/standards/si/si.html>

[20] Michael Everson, 1996. *Report of the Sinhala Standards*, ISO/IEC JTC1/SC2/WG2 N1473R.

[21] Gunasekara, Abraham Mendis, 1891. *A comprehensive Sinhalese Language*, Asian Educational Services, New Delhi.

[22] SLS 1134:1996. *Sri Lanka Standard SLS 1134:1996-Sinhala Character Code for Information Interchange*, SLSI publication.

[23] Lloyd Anderson, Ken Whistler, Peter Lofting, Rick McGowan (Contributors), 1992. *Draft Proposals Unicode Technical Report #2* Unicode Inc.,

[24] Hugh McGregor Ross, 1996. *Sinhala proposal*, ISO/IEC JTC1/SC2/WG2 N1376.

[25] Michael Everson, Takayuki Sato, Kohji Shibano, Disanayaka J. B., Nandasara S. T., Johan van Wingen & Glenn Adams, 1997. *Minutes of Sinhala Ad-Hoc Committee, Doc. # 1613, ISO Meeting No. 33, Iraklion, Crete, Greece*.

[26] Naito Eisuke, 1998. *Progress Report of the MLIT Project*, AFSIT-12, Ha Noi, Vietnam.

[27] Takayuki K. Sato, 1998. *Status of Cooperative Activities for the Missing Characters*, MLIT Secretariat, CICC, Japan.

- [28] Unicode Standard 3.0, The, 1998. (www.unicode.org), Addison-Wesley Pub Co., ISBN 02001616335
- [29] SLS 1134:2004. *Sri Lanka Standard SLS 1134:2004-Sinhala Character Code for Information Interchange*, SLSI publication.
- [30] Wijayananda Jayaweera, *Kothmale Community Radio/Internet Project: Expanding the Knowledge Base*, UNESCO, June, 2001
- [31] Nandasara, S. T., *GII/GIS for Equal Language Opportunity*, MLIT-3: October 6 - 7, Ha Noi, Vietnam, 1998

Bibliography

- Daniel J. T. K. & Hedlund R. E. *Carey's Obligatio and India's Renaissance*, Council of Serampore College, Serampore, West Bengali, 1993.
- De Silva, M. W. Sugathapala. *Some Consequences of Diglossia*, 1982.
- Edward Miller, *Ancient Inscriptions in Ceylon*, PLATES, Trubner & Co. Ludgate Hill, London, 1883.
- Jayasiri Lankage, *Evolution of Sinhala Scripts*, S. Godage & Pvt Ltd, Colombo, Sri Lanka, 1996.
- John Devy, *An Account of the interior of Ceylon*, First published in 1812. New Delhi, AES, 1990.
- Paranavitana, S, *Inscriptions of Ceylon*, Vol. I, 1970.
- Piyadasa T. G., *Libraries in Sri Lanka, Their Origin and History from Ancient Times to the Present Time*, Sri Satguru Publication, India, 1985.
- Wilfred A. Beeching, *Century of Typewriter*, Heinemann, London, 1974.