

Sinhala as a living language in the Digital (Internet) Age

Harsha Wijayawardhana B.Sc. in Biochemistry (Miami) , CITP (UK), FBCS(UK)
Chairman/member Local Language Working Group (LLWG) of ICTA
Member-Advisory Committee of SDCI
COO/CTO Theekshana R & D (Affiliated Company of UoC)
Board Member, BoC and LK Domain registry

Topics covered

- Evolution of Sinhala Language and its Script
 - Special Reference to Pali, Sanskrit, and Tamil
- Influence of Buddhism on Sinhala Language and its role
- Script vs. Language
- Encoding of Sinhala Script in Unicode and Standardization of SLS 1134
- Encoding of Sinhala Numerals
- DNA Research on Sinhala People
- Sinhala on the Internet-Social Media and Tools such as OCR, Speech to Text, Translator tool development
- Digitization and automation of Sinhala Dictionary Compilation Institute

Landmark events of the Sinhala Language

- Evidence for writing, record keeping going back before Vijaya, founder of Sinhala Race –500 BC (Pottery Sherds Anuradhapura and Thissamaharma) Dr. Shiran Deraniyagala, Dr. Mahadeva, Prof. Raj Somadeva.
- 3rd Century BC, arrival of Arahant Mahinda and translation of commentaries to Elu-Sihala.
- 1st Century BC, writing down or recording of Tripitaka during the reign of Valagamba in Sri Lanka
- In 5th Century AD, Buddhagosha Thera from Bihar (a Brahmin) translating Sihala Atuwa or commentaries to Pali (Magadhi)
- Recording of Sri Lankan history –Dipavansa, Mahavamsa (Mahanama thera)
- Sinhala Buddhist monks involvement in spreading Buddhism throughout East Asia.
- In mid 1700s AD, restoring of Upasamada through the assistance of Thailand (Asarana Sarana Saranakara Sanaga Raja Thero and Upali Thero from Thailand)
- Establishment of Two Pirivenas Vidyodaya and Vidyalankara in the late 1800s
- Encoding of Sinhala in Unicode in 1996 and Implementation of Sinhala on Computers between 2002-to-date.

Elu and Pali

- Quote from a paper written by the Sri Lankan Scholar Monk, Walpola Rahula Thero on Pali “It is curious that the language which call Pali- the Language of Theravada Buddhist Literature- is not found anywhere in India either in ancient inscriptions or in any literary work. The Language of some of the edicts of Asoka, though similar to it, is not Pali. Pali was first found in Sri Lanka and from there it spread over to other Theravada countries with the Tripitaka and commentaries. Although it was found originally in Sri Lanka, nobody believes that Pali originated in in the “Island of the Sinhala” (Sihaladipa).

(Humor in Pali Literature and Other Essays by Walpola Rahula, published by Walpola Rahula Foundation Trust)

Buddhist influence ...

- Rys Davids- Author of Buddhist India calls Elu – Prakrit of Ceylon
- Rys Davids pays great homage to Sinhala writers who recorded Theravada canon alluding to Sinhala people that they had treated Tripitaka as closed or closed canon.
- Rys Davids –”In Ceylon they had been much worked up, polished, elucidated, systematized. From Several works now accessible we know fairly well the tone and manner of these later---and, as they must have seemed to Ceylon scholars, clearer, fuller---statements of the old ideas. In no single instance yet discovered has this later tone and manner found its way into the canonical books”

Sinhala

- Sinhala belongs to Indo-European Language family and sub family of Indo- Arya.
- Closest Relatives of Sinhala Language are from North Indian family such as Hindi, Urdu, Bengali, Orya, etc.
- Europe has several of Indo-European Sub-families such as German, Romanic, and East Slavic.
- English, German and Swedish belong to German Sub Family whereas French, Portuguese, Spanish and Italian belong to Romanic Languages.



Written Script vs. Language

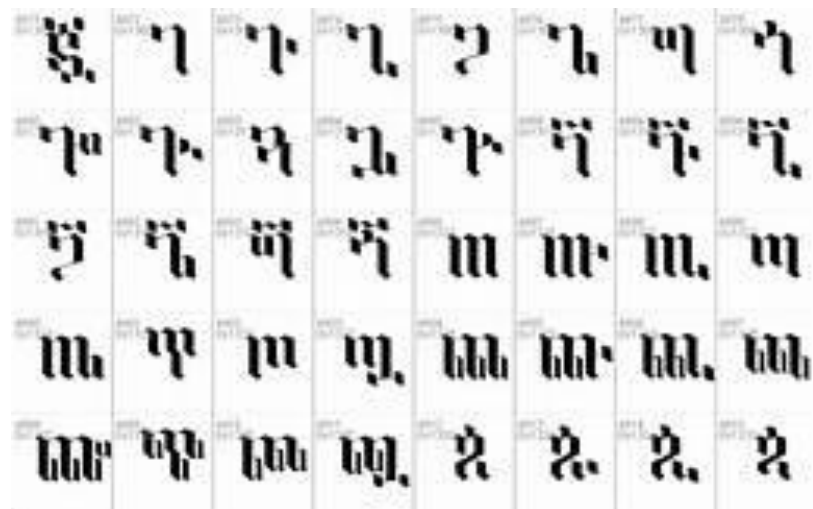
- Many spoken Languages in the world and some without a written scripts.
- At present, world has about fifty written scripts. It is believed that scripts derived from Phoenicians who were northern part of Middle East or from present day Israel.
- It is believed that the modern

Alphabet came into existence
between 1800 -1900 BC.

- Modern Indian and East Asian scripts derive from Brahmi Scripts and they belong to Abugida Scripts.

[illegible]

Written Scripts



Ethiopic

ก ข ค ฅ ง จ ฉ ช
 ฌ ฌ ญ ฎ ฏ ฐ ฑ
 ฒ ด ต ถ ท ธ น บ ป
 ผ ฝ พ ฟ ภ ม ย ร ล
 ว ศ ษ ส ห ฬ อ ฮ

www.pudthai.com

Thai Script

འབྲུག་ཡུལ་

ཨ་ཁུ་ལཱ་

Dzongkha

skā	ykka	ccha	jja	ṭṭa	ḍḍa	ṣṣa
pta	bda	sna	ppa	kma	psa	kka
kta	ṇka	ṇṇa	ṇṇa	ṇca	ṇṭa	kṣa
ṇṇa	ṇṇa	ṇṇa	ṇṇa	ṇṇa	ṇṇa	ṇṇa
ṇṇa	ṇṇa	ṇṇa	ṇṇa	ṇṇa	ṇṇa	ṇṇa

Consonant diacritics

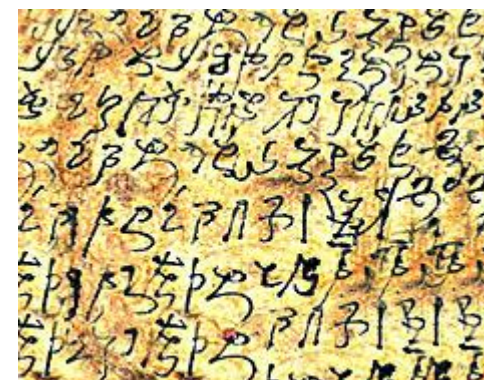
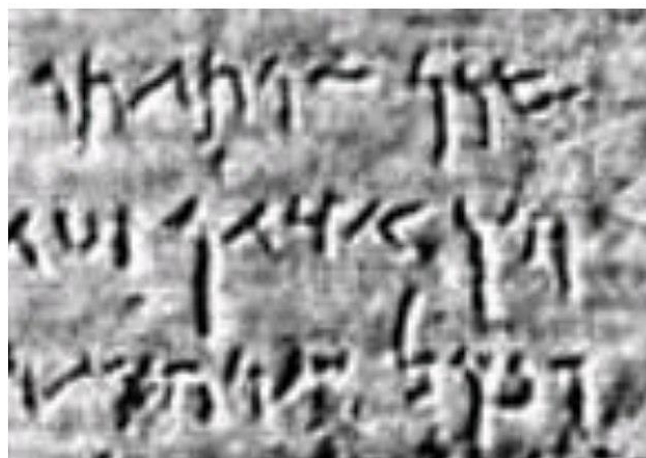
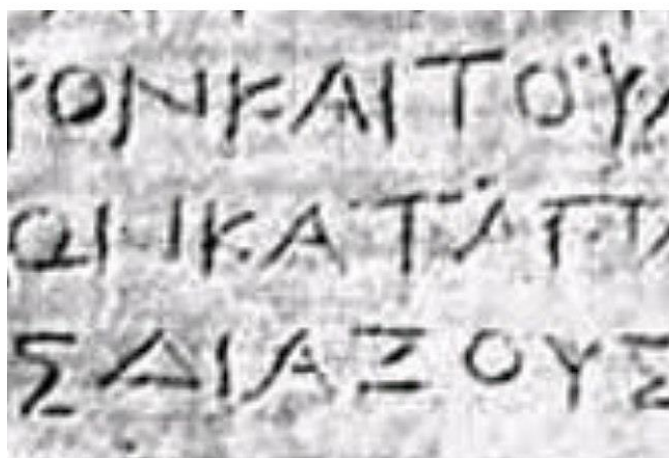
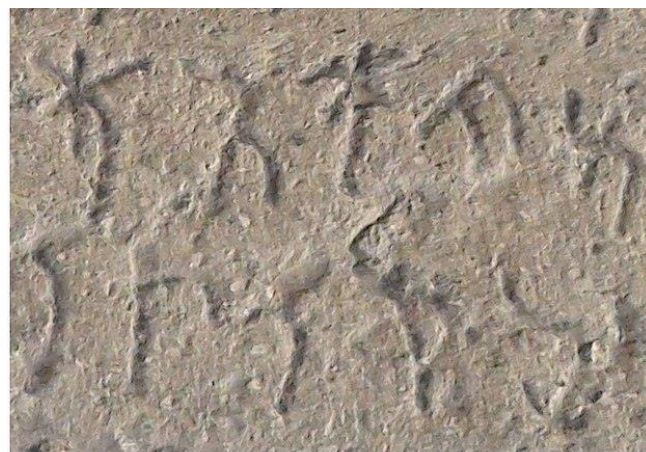
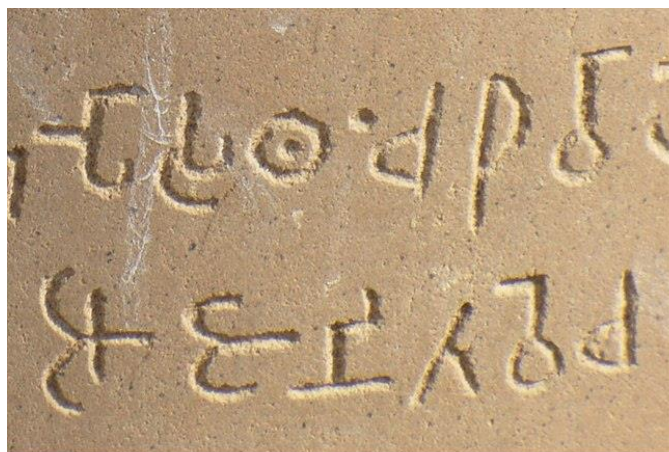
nya	pra	śva
-----	-----	-----

Malayalam

Brahmi

- All these Indian, East Asian Scripts derived from Brahmi.
- Although it was once believed that Emperor Asoka used Brahmi as his regal script, it has discovered that oldest Brahmi is found in Sri Lanka and Southern India.
- Brahmi is usually written from right to left. In Sri Lanka, eighteen (18) Rock inscriptions have discovered with Brahmi written from right to left dating to 3rd century BC.
- In North Eastern part of Asoka's Empire, his edicts were written in Prakrit using Kharoshthi Script. Kharoshthi was written from right to left.

Brahmi and Kharosthi



Kharosthi Script



Brahmi

Brahmi Script

- It was earlier postulated that Brahmi derived from Aramaic.
- Another hypothesis is Brahmi derived from Indus Valley script.



Sinhala and other Indic Scripts

- Sinhala Script derived from Southern Brahmi Script.
- Subsequently, Sinhala Script was influenced by Pallava Grantha.
- Eastern Scripts such as Khmer, Myanmar, etc. influenced by Pallava Script.
- Sinhala has two Vowels අ and ඞ.
- These two vowels are found in East Asian Scripts.



Thai and Khmer



Devanagari	क	ख	ख	ग	ग	घ	ङ	च	छ	ज	ज	झ	ञ	ट	ट	ठ	ड	ढ	ण	त	त
Thai	ก	ข	ฃ	ค	ฅ	ฆ	ง	จ	ฉ	ช	ฌ	ญ	ฎ	ฏ	ฐ	ฑ	ฒ	ณ	ด	ต	ท
Burmese	က	ခ		ဂ		ဃ	င	စ	ဆ	ဇ		ည	ဉ		တ	ထ	ဓ	ဋ	ဌ	ဍ	ဎ
Khün	က	ခ	ၵ	ဂ	ၵ	ဃ	င	စ	ဆ	ဇ	ၵ	ည	ဉ		တ	ထ	ဓ	ဋ	ဌ	ဍ	ဎ
Lanna	က	ခ	ၵ	ဂ	ၵ	ဃ	င	စ	ဆ	ဇ	ၵ	ည	ဉ		တ	ထ	ဓ	ဋ	ဌ	ဍ	ဎ
Khmer	ក	ខ		ក		យ	ង	ច	ឆ	ជ		ឈ	ញ		ដ	ប	ឌ	ឍ	ណ		ត

Devanagari	थ	द	ध	न	प	प	फ	फ	ब	फ	भ	म	य	र	ल	ळ	व	श	ष	स	ह	ह	अ
Thai	ถ	ด	ธ	น	ป	ป	ผ	ฝ	พ	ฟ	ภ	ม	ย	ร	ล	ฬ	ว	ศ	ษ	ส	ห	ฮ	อ
Burmese	ထ	ဒ	ဓ	န	ပ	ပ	ဖ	ဖ	ဘ	ဖ	မ	ယ	ရ	လ	လ	လ	ဝ			သ	ဟ	အ	
Khün	ထ	ဒ	ဓ	န	ပ	ပ	ဖ	ဖ	ဘ	ဖ	မ	ယ	ရ	လ	လ	လ	ဝ	ဇ	သ	ဟ	ဟ	အ	
Lanna	ထ	ဒ	ဓ	န	ပ	ပ	ဖ	ဖ	ဘ	ဖ	မ	ယ	ရ	လ	လ	လ	ဝ	ဇ	သ	ဟ	ဟ	အ	
Khmer	ថ	ទ	ធ	ន	ប	ប	ផ		ព		ភ	ម	យ	រ	ល	វ	ឡ	រ	គ	ឝ	ស	ហ	អ

Unicode Version 13.1

The Unicode Standard, Version 13.1

unicode.org/charts/PDF/U0D80.pdf

Apps | How to Fix Msvcrt11... | GitHub - in28minut... | Test-Tamil-MFA | TEST_ENG_MFA | TEST-SIN-MFA | Other bookmarks

The Unicode Standard, Version 13.0 | 2 / 4 | 49% | [Navigation icons]

1

2

3

0D80 Sinhalā 0DFF

	0D80	0D81	0D82	0D83	0D84	0D85	0D86	0D87	0D88	0D89	0D8A	0D8B	0D8C	0D8D	0D8E	0D8F
0	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
1	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
2	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
3	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
4	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
5	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
6	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
7	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
8	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
9	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
A	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
B	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
C	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
D	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
E	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ
F	ආ	භ	ඈ	ඉ	ඊ	උ	ඌ	ඍ	ඎ	ඏ	ඐ	එ	ඒ	උ	ඌ	ඍ

The Unicode Standard 13.0. Copyright © 1991–2020 Unicode, Inc. All rights reserved.

Type here to search

10:13 PM 3/1/2021

Sinhala Archaic Numbers in Supplementary Multilingual Plane (SMP)

ISO/IEC 10646:2011 FDIS

Not secure | unicode.org/L2/L2011/11220-n3968-pdam10-all.pdf

Apps | How to Fix Msvcrt1... | GitHub - in28minut... | Test-Tamil-MFA | TEST_ENG_MFA | TEST-SIN-MFA | Other bookmarks

ISO/IEC 10646:2011 FDIS

19 / 33 | 50%

18

19

20

Proposed Draft Amendment (PDAM) 1
111E0

Sinhala Archaic Numbers
111FF

This number system is also known as Sinhala Bhikkam. This number system does not have a zero place holder concept unlike the Sinhala astronomical numbers, Sinhala Laka Bhikkam, provided in the UDHR-HEWP range.

Historical digits
These digits are not used with a zero

111E1	111E2	111E3	111E4	111E5	111E6	111E7	111E8	111E9
අ	ආ	ඇ	ඈ	ඉ	ඊ	උ	ඌ	ඍ

Historical numbers

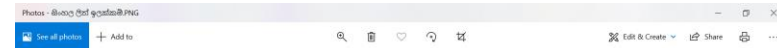
111EA	111EB	111EC	111ED	111EE	111EF	111F0	111F1	111F2	111F3	111F4
දස	විස	සිය	සිය	සිය	සිය	සිය	සිය	සිය	සිය	සිය

© ISO/IEC 2011 – All rights reserved

10:17 PM
3/1/2021

Sinhala Numerals

- Although five ways of writing numbers found, only two have been encoded.
- The Sinhala Illakkam and Lith Illakkam have evolved from Brahimi numerals.
- Lith Illakkam has a zero but not the Sinhala Illakkam.



෦ ෧ ෨ ෩ ෪ ෫ ෬ ෭ ෮ ෯



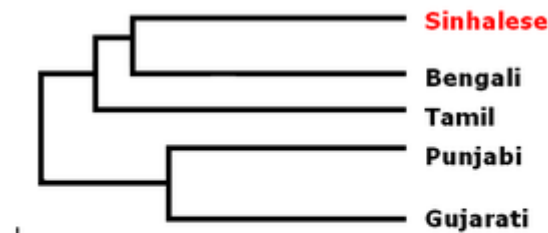
Language of Vedda (Aborigines of Sri Lanka)

- According to Sri Lankan history, Vaddas are descendants of the two children of King Vijaya, the North Indian Prince, and the local princess Quveni.
- Some groups of Veddas still live as hunter-gathers on the remote parts of the island.
- Many scholars who had studied their language conclude that it is a creole version of Sinhala.
- It is most probable that Veddas descend from the original settlers who arrived in Sri Lanka many thousand years back and are living descendants of Balangoda Man, who lived thirty-five thousand years ago.
- The language of Vedda, the creole version of Sinhala, comes under UNESCO classification as an endangered language in Asia.

DNA Research on Sinhala people

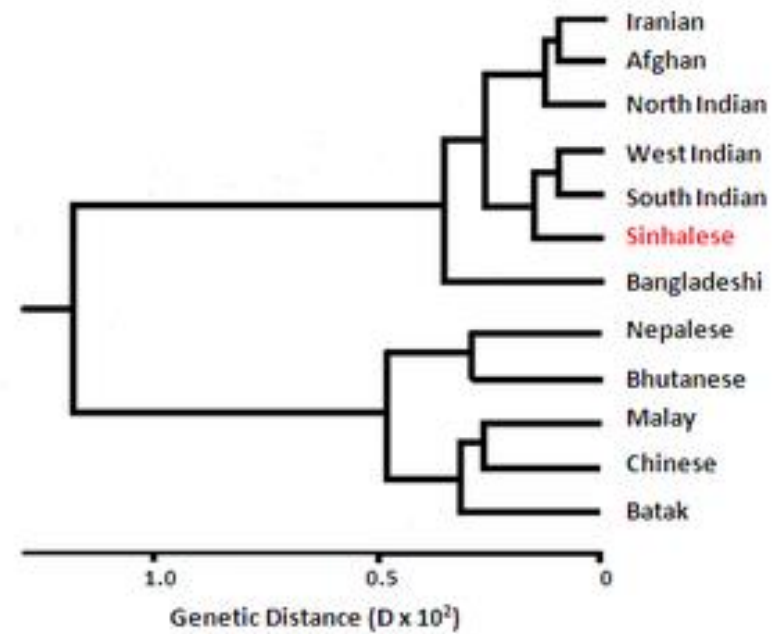
- Sinhala people are more closely related to Bengalis than other South Asians
- Sinhala people carry R1a1a Indo European marker of Y chromosomal Haplo group
- A study in 2007 found similar frequencies of the allele [HLA-A*02](#) in Sinhala (7.4%) and North Indian subjects (6.7%). HLA-A*02

is a rare allele which has a relatively high frequency in North Indian populations and is considered to be a novel allele among the North Indian population. This suggests possible North Indian origin of the Sinhala.



Adapted from: Mastana S (2007) "Molecular Anthropology: Population and Forensic Genetic Applications.", The Anthropologist, 3:373-383,

DNA Sinhala cont.



Adapted from: Raychoudhury AK, Nei M (1985) Genetic relationships between Indians and their neighboring populations. Hum Hered 35:201-206

Sinhala on the Social Media

- Although Sinhala is not classified as an endangered language yet, Sinhala could become extinct within the century, provided if people were reluctant to use Sinhala for day-to-day communication.
- Digitally, Sinhala has become a very viable language. Therefore, the danger of the disappearance of the Sinhala Script as a living script and language from the world is minimal.
- The latest observation is that Sinhala is widely used on Social Media and a new genre of literature and a writing style seem to emerge through the new digital media.

Sinhala on the Social Media cont ...

- The government must encourage new Translation tools from Sinhala to Tamil and English and vice versa using Speech to Text and Text to Speech.
- If machine translation tools were to be available for English to Sinhala, students who are not well versed in English could learn using the content available on the Internet.

Digitization of SDCI

- Developing an effective Website for the SDCI with a payment Gateway
- Mobile Apps development to popularize Sinhala words, meanings, etc.
- Digitization and Automation of Dictionary Compilation process.
- Use of modern technology such as Book Scanners with Optical Character Recognition (OCR) Software, etc.
- Building Sinhala Corpus ---Research on Sinhala use.

- Thank you ...

Any Questions?

Questions and Comments ---email wijayawardhana@gmail.com

Mobile No: 0777 872 642