ශ්‍රී ලංකා ප්‍රමිතිය 1134 : 2004
SRI LANKA STANDARD 1134 : 2004
UDC 003.035 : 003.336

තොරතුරු හුවමාරුව සඳහා වූ සිංහල අක්ෂර කේතය (දෙවන ප්‍රතිශෝධනය)

SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE (SECOND REVISION)

SRI LANKA STANDARDS INSTITUTION
SINHALA CHARACTER CODE FOR INFORMATION INTERCHANGE
(SECOND REVISION)

SLS 1134 : 2004

Gr. 11

SIR LANKA STANDARDS INSTITUTION
No. 17, Victoria Place
Elvitigala Mawatha
Colombo 08
SRI LANKA

Sri Lanka Standards are subject to periodical revision in order to accommodate the progress made by industry. Suggestions for improvement will be recorded and brought to the notice of the Committees to which the revisions are entrusted.

This standard does not purport to include all the necessary provisions of a contract.

CONTENTS

Sri Lanka Standard
SINHALA CHARACTER CODE FOR INFORMATIONINTERCHANGE
(Second Revision)

FOREWORD

This standard was approved by the Sectoral Committee on Information Technology and was authorised for adoption and publication as a Sri Lanka Standard by the Council of the Sri Lanka Standards Institution on 2004-11-10.

The International Organisation for Standardisation accepted **SLS 1134** Sinhala Character Code for Information Interchange for inclusion in **ISO/IEC 10646-1** with the modifications effected in the first revision (**SLS 1134:2001**).

This is the second revision of **SLS 1134.** This revision provides specifications for the code sequences and keyboard sequences. It also provides a revised keyboard, based on the layout in the original version of this standard, which in turn is based on the Wijesekara typewriter keyboard. This revision retains compliance with **ISO/IEC 10646-1**.

Symbols used in the Sinhala language are coded using 128 cells in the half page plane reserved for Sinhala characters in **ISO/IEC 10646.** Each cell or position given in Figure **1** of the standard represents one character.

An effort has been made to preserve the alphabetical order of the Sinhala language to a great extent. However, specific collation algorithms (not specified in this standard) are required to correctly collate text encoded in this code.

In the preparation of this standard the valuable assistance obtained from the following publications is gratefully acknowledged.

ISO/IEC 10646-1:1993 Information Technology – Universal Multiple Octet Coded Character Set.

The Unicode Consortium: **The Unicode Standard Version 4.0, 2004**.

The assistance provided by the Council for Information Technology (CINTEC) and the Information and Communication Technology Agency of Sri Lanka (ICTA) in the preparation of this standard is also gratefully acknowledged.

1 SCOPE

This standard provides a coding of the set of Sinhala characters for use in computer and communication media. This standard character code set specifies a 7-bit code table (out of 16 bits) which may be used in line with the requirements outlined by the International Organisation for Standardisation (ISO).

In addition to storage, retrieval and machine to machine communication in Sinhala, it also includes provisions to co-exist with other languages as specified in **ISO/IEC 10646.**

This code set is able to represent contemporary and historical Sinhala writings.

This standard defines codes for the vowels, consonants, semi-consonants, signs, and punctuation in the language. Some formations of the language are not represented by individual codes, but are constructed as sequences of codes. For example, many characters are formed by a consonant followed by a consonant modifier.

In designing this code set efforts were taken to retain the ability to incorporate future developments in the language.

It does not include editorial characters, abbreviations, subscribes, superscribes and punctuations, but in the keyboard layout, keys are provided for Indo-Arabic numerals, symbols and punctuation.

NOTE: *Codes are not provided in the code set for distinct formations in the language for the Repaya, Yansaya, and Rakaaraansaya. These shapes are generated using the sequences given in* ***5****.*

2 DEFINITIONS

For the purpose of this standard the following definitions shall apply:

2.1 base character: A character which may stand alone, or optionally combine with one or more combining characters.

2.2 base letter: A symbol from which letters are formed.

2.3 character: A unit of information used for the organisation, control, or representation of textual data.

2.3 code table: A table showing the characters allocated to the cells in the code.

2.4 combining character: A character which does not stand alone, but combines with another character.

2.5 combining character sequence: A sequence of characters, starting with a base character and zero or more combining characters, which represent a letter.

2.6 composite character: A character which is equivalent to a sequence of one or more other characters.

2.7 conjunct letter: Two or more letters joined together. In addition to the combination of a consonant with the yansaya, rakaaraansaya and repaya, pairs of consonants may form conjunct letters.

2.8 consonant: A letter representing a speech sound in which the breath is at least partly obstructed and which may combine with a vowel to form a syllable.

2.9 letter: A symbol representing a simple or compound sound used in speech. It may comprise a base letter, or a base letter together with one or more strokes.

2.10 non-vocalic stroke: A graphic symbol associated with a base letter to indicate a consonant to be associated with that letter.

2.11 pure consonant: A consonant without an associated vowel, i.e., with an al-lakuna.

2.12 semi-consonant: A consonant that does not enjoy all the privileges of normal consonants and to be combined with the preceding vowel.

2.13 sign: One or more strokes.

2.14 stroke: A graphic symbol which modifies a base letter.

2.15 vocalic stroke: A graphic symbol associated with a base letter to indicate the presence or absence of a vowel associated with that letter.

2.16 vowel: A letter representing a speech sound made with the vibration of the vocal cords, but without audible obstruction.

3 DESCRIPTION OF THE SINHALA LANGUAGE

Sinhala is a member of the Indo-Aryan family of languages and the script bears close structural resemblance to Thai and Malayalam scripts. The Sinhala writing system is a syllabary derived from the ancient North Indian script, Brahmi, and subsequently influenced by the Pallawa Grantha script of South India. The modern script used in writing Sinhala is unique to this language.

Sinhala differs from all other Indo-Aryan languages in that it contains a pair of vowel sounds that are unique to it. These are the two vowel sounds that are similar to the two vowel sounds that occur at the beginning of the English words, *at* and *ant.* The vowel sound in *at* is short, and the vowel sound in *ant* is long. The Sinhala alphabet has a pair of letters to represent these two sounds.

Short vowel: ඇ - ae
Long vowel: ඈ - aee

Another feature that distinguishes Sinhala from its sister Indo-Aryan languages is the presence of a set of five nasal sounds known as “half nasal” or “prenasalized stops”. These sounds as represented in modern Sinhala writing and their romanised notation are as follows:

ඟ (nng), ඦ (nyj), ඬ (nndd), ඳ (nd), ඹ (mb)

The Sinhala alphabet (as defined below) consists of 61 letters: 18 vowels, 41 consonants and 2 semi-consonants.

<Sinhala alphabet>:: = <Vowels ><Consonants>< Semi-consonants>

These symbols represent 40 sounds: 14 vowel sounds and 26 consonant sounds.

The 61 letter symbols are given below together with their romanised representations.

3.1 Vowels

The 18 vowels, unlike consonants, are used only at the beginning of words.
They are as follows:

අ a	ආ aa	ඇ ae	ඈ aee
ඉ i	ඊ ii	උ u	ඌ uu
ඍ ṛ	ඎ ṛṛ	ඏ ḷ	ඐ ḷḷ
එ e	ඒ ee	ඓ ai	
ඔ o	ඕ oo	ඖ au	

NOTES :

1. *The letters ඇ and ඈ which are based on the symbol අ are unique to the Sinhala language and have been in use since the 7th century.*

2. *ඏ (ilu) and ඐ (iluu) do not occur in present usage but are included in the code set for completeness of the code. They are not included in Tables 1 - 3.*

3. *ඎ also does not occur in present usage, but its corresponding vowel sign, ෘා is used; for example, ශාස්තෘන්.*

3.2 Consonants

The Sinhala alphabet possesses 41 consonants as shown below with the romanised notations as used in Unicode.

ක	ka	ඛ	kha	ග	ga	ඝ	gha	ඞ	nga	ඟ	nnga		
ච	ca	ඡ	cha	ජ	ja	ඣ	jha	ඤ	nya	ඦ	jnya	ඵ	nyja
ට	tta	ඨ	ttha	ඩ	dda	ඪ	ddha	ණ	nna	ඬ	nndda		
ත	ta	ථ	tha	ද	da	ධ	dha	න	na	ඳ	nda		
ප	pa	ඵ	pha	බ	ba	භ	bha	ම	ma	ඹ	mba		
ය	ya	ර	ra	ල	la	ව	va						
ශ	sha	ෂ	ssa	ස	sa	හ	ha	ළ	lla	ෆ	fa		

NOTES:

1. *The consonant ඵ (nyja) is included although it is not found in contemporary writing.*

2. *The letters ඤ and ඳ are identical in sound only in the initial position of a word for example ඤාණ and ඳානා. They are not identical in non-initial positions, where ඳ behaves as a combination of two consonant sounds, for example ප්‍රඳා.*

3.3 Semi-consonants

"ං" and "ඃ" are the two semi-consonants available in the alphabet. Semi-consonants may appear only following a vowel or a consonant with an implicit or explicit vowel.

NOTES:

1. *The corresponding phonetic notations of semi-consonants are:*
   ං *ṃ* ඃ *ḥ*
2. *These two characters have been placed at the beginning of the code set to facilitate collation.*

3.4 Strokes

Strokes (also known as modifiers) are graphical symbols used in conjunction with consonants. They are also used in writing some vowels (e.g. ඇ, ෆ, ඓ). The strokes of the Sinhala script occur in two different forms; i.e., vocalic strokes and non-vocalic strokes.

<Stroke> :: = <Vocalic stroke>, < Non-vocalic stroke>
<Vocalic stroke> :: = ැ, ෑ, ි, ී, ු, ූ, ෘ, ෲ, ෟ, ෳ
<Non-Vocalic stroke> :: = ්, ්‍ර, ෘ

Unlike in English, a stroke may be positioned on any of the four sides of the base letter. These can be classified as follows:

<Stroke> :: = <Left stroke>, <Right stroke>, <Upper stroke>, <Lower stroke>
<Left stroke> :: = ෙ
<Right stroke> :: = ා, ෑ, ෟ, ෳ
<Upper stroke> :: = ැ, ි, ී, ෘ
<Lower stroke> :: = ු, ූ, ෲ

The strokes, their names, and the vowels represented by them, are given in Table **1**.

3.5 Letters

In Sinhala, a letter may be formed by a vowel alone, a consonant alone, or a consonant with one or more associated strokes. It may optionally include a semi-consonant.

Seventeen combinations of vocalic strokes with a consonant are used in Sinhala. These are listed in Table **2**.

The non-vocalic strokes ්‍ය and ්‍ර represent the letter ය and ර respectively, following a pure consonant. They may appear alone, or together with vocalic strokes. If so, the vowel applies to the ය or ර and not the initial consonant. The valid combinations of vocalic strokes with the yansaya and rakaaraansaya (7 and 12 combinations respectively) are shown in Table **3**.

A semi-consonant, ං or ඃ, may appear after any of the letters in lines 2 to 17 in Table **2** and any of the letters in Table **3** (i.e., after any letter except a pure consonant), yielding a total of 109 possible letters based on a given consonant.

TABLE 1 – Strokes, their names and vowel representation

Sl. No. (1)	Stroke (2)	Name (3)	Vowel repre-sentation (4)
Vocalic strokes			
1	්	Sinhala al-lakuna l	-
1a	්*	Sinhala al-lakuna 2	-
2	ා	Sinhala aela-pilla	ආ
3	ැ	Sinhala ketti aeda-pilla	ඇ
4	ෑ	Sinhala diga aeda-pilla	ඈ
5	ි	Sinhala ketti is-pilla	ඉ
6	ී	Sinhala diga is-pilla	ඊ
7	ු	Sinhala ketti paa-pilla 1	උ
7a	කු*	Sinhala ketti paa-pilla 2	උ
8	ූ	Sinhala diga paa-pilla 1	ඌ
8a	කූ*	Sinhala diga paa-pilla 2	ඌ
9	a	Sinhala gaetta-pilla	ඍ
10	ෙ	Sinhala kombuva	එ
11	ෘ	Sinhala gayanukitta	ඔ
Non Vocalic strokes			
12	්‍ය	Sinhala yansaya	refer note 3
13	් ර	Sinhala rakaaraansaya	refer note 3
14	් ර්	Sinhala repaya	refer note 3

*These strokes are shown with an associated character.

NOTES:

1. *The **al-lakuna** removes the implicit vowel associated with a consonant, forming a pure consonant.*

2. *The shape of a stroke is dependant on the associated consonant, as shown in lines 1, 7 and 8 above.*

3. *The non-vocalic stroke Sinhala **yansaya** symbolizes ය when preceded by a pure consonant, e.g.: ක් + ය = ක්‍ය. The non-vocalic stroke Sinhala **rakaaraansaya** symbolises ර when preceded by a pure consonant, e.g.: ක් + ර = ක්‍ර. The **repaya** symbolises a ර් preceding a consonant, e.g.: ක + ර් + ම = කර්ම.*

4. *The vocalic stroke ෘ **diga gayanukitta** corresponding to the vowel ඖ is not presently used in Sinhala and is not considered further.*

TABLE 2 - Combination of the consonant ක (k) with vocalic strokes

Sl No. (1)	Character (2)	Phonetic Notation (3)
1	ක්	k
2	(ක් + අ) = ක	ka
3	(ක් + ආ) = කා	kaa
4	(ක් + ඇ) = කැ	kae
5	(ක් + ඈ) = කෑ	kaee
6	(ක් + ඉ) = කි	ki
7	(ක් + ඊ) = කී	kii
8	(ක් + උ) = කු	ku
9	(ක් + ඌ) = කූ	kuu
10	(ක් + ඍ) = කෘ	kṛ
11	(ක් + ඎ) = කෲ	kṛṛ
12	(ක් + එ) = කෙ	ke
13	(ක් + ඒ) = කේ	kee
14	(ක් + ඓ) = කෛ	kai
15	(ක් + ඔ) = කො	ko
16	(ක් + ඕ) = කෝ	koo
17	(ක් + ඖ) = කෞ	kau

TABLE 3 - Combination of the consonant ක (k) with vocalic and non-vocalic strokes

Sl No. (1)	Character (2)	Romanied Notation (3)	Sl No. (1)	Character (2)	Romanised Notation (3)
1	කය	kya	9	ක්‍ර	kra
2	කයා	kyaa	10	ක්‍රා	kraa
3	කයු	kyu	11	ක්‍රැ	krae
4	කයූ	kyuu	12	ක්‍රෑ	kraee
5	කෙය	kye	13	ක්‍රි	kri
6	කෙය්	kyee	14	ක්‍රී	krii
7	කොය	kyo	15	ක්‍රෙ	kre
8	කොය්	kyoo	16	ක්‍රේ	kree
			17	ක්‍රෛ	krai
			18	ක්‍රො	kro
			19	ක්‍රෝ	kroo
			20	ක්‍රෞ	krou

NOTES:

1. *The yansaya is not used following the letter ර. e.g.: the spelling කාර්‍ය is incorrect.*

2. *Not all combinations are valid with all consonants. E.g., the consonant ඬ is never combined with a vowel, but appears as ඬ. However, we do not list all such cases.*

3 *The **repaya** is not included in this table.*

3.5.1 *Repaya*

The repaya is an abbreviation for the letter ර් preceding a consonant. Use of the repaya is optional. e.g.: කර්‍ම or කර්ම are both valid.

3.5.2 *Conjunct Letters (බැඳි අකුරු)*

A conjunct letter is formed by joining another letter to a pure consonant (e.g. ක් + ෂ = ක්‍ෂ). A conjunct letter may be modified by vocalic strokes (e.g. ක්‍ෂී). The strokes yansaya (්‍ය) and rakaaraansaya (්‍ර) are shortened forms of the letters ය and ර respectively, used when forming conjunct letters. These strokes may also be appended to a conjunct letter, e.g. ඳ.

3.5.3 *Touching Letters*

Classical and Buddhist texts, in both the Sinhala and Pali languages, use the convention of representing a pure consonant followed by another consonant by omitting the al-lakuna and writing the consonants touching each other, e.g. කඤ.This representation is similar, but not identical to, conjunct letters. Some pairs of letters may be written in all three forms (normal, "touching" and conjunct), e.g., ද්ධ, ඬ and ඹ.

4 CHARACTER ENCODING

The encoding of the Sinhala character set into 128 cells of a 16-bit code space is shown in Figure **1**. This encoding uses hexadecimal codes in the range 0D80 to 0DFF.

The table comprises codes for the semi-consonants (0D82-0D83), vowels (0D85-0D96), consonants (0D9A-0DC6), the al-lakuna (0DCA), vowel signs (0DCF-0DF3) and punctuation mark - kundaliya (0DF4). Some vowel signs are composite characters which represent two or more vocalic strokes. However, each vowel sign corresponds to a vowel. The unused positions in the range shall not be used.

Vowels, consonants and the punctuation mark are base characters, and may stand alone. The al-lakuna and vowel signs are combining characters, and follow the code of a consonant. The semi-consonants are combining characters, and follow a vowel, consonant, or vowel sign.

Descriptions of the codes are given in Table **4**.

NOTES:

1. *The Sinhala character encoding represented in Figure **1** is identical to ISO 10646-1 and Unicode Version 4.*

2. *Specific collation algorithms (not specified herein) are needed for sorting text stored using this character encoding.*

4.1 Codes for numerals

As Sinhala numerals are not used at present, codes are not provided for Sinhala numerals.

4.2 Code for Sinhala punctuation

The code 0DF4 represents the **kundaliya.**

NOTE : *The kundaliya, which is a punctuation mark unique to Sinhala writing, is sometimes used to conclude a paragraph.*

4.3 Codes not specified herein

This standard does not specify codes for the numerals, other punctuation marks and symbols. These are specified in **ISO/IEC 10646**. The yansaya, rakaaraansaya and repaya are represented by code sequences, and not by individual codes.

The codes for zero-width joiner (ZWJ), which is used to form conjunct letters, and zero-width non-joiner (ZWNJ), are not listed in **Figure 1**, but have the values 200D and 200C respectively. The ZWNJ is not used in encoding Sinhala, but implementations should ignore any ZWNJ characters encountered when processing Sinhala text.

The non-breaking space character (NBSP) has the value 00A0.

The code for INV is not specified in this standard (see **5.10**).

4.4 Codes reserved for future developments

Three codes each after the sets of vowels (0D97-0D99) and consonants (0DC7-0DC9) are left unassigned to accommodate future enhancements of the language.

5 CODE SEQUENCES

Each Sinhala letter (e.g., කේ) is represented by sequence of characters in Figure **1**. A letter may be a vowel (e.g. ඒ), a consonant (e.g. ක), a consonant followed by an al-lakuna (i.e., a pure consonant) (e.g. ක්), a consonant with a vowel sign (e.g. කා), one of the above (except a pure consonant) followed by a semi-consonant (e.g. අං), or a conjunct letter (e.g. ක්‍ෂ) optionally followed by an al-lakuna, vowel sign and/or a semi-consonant.

5.1 Vowels

Each vowel is represented by one character in the range 0D85-0D96.
e.g. අ = 0D85, ඍ = 0D8D

NOTE: *A vowel such as ඇ should* ***not*** *be represented as a character sequence such as 0D85 0DCF.*

5.2 Consonants

Each consonant is represented by one character in the range 0D9A – 0DC9.
e.g. ක = 0D9A, ඤ = 0DA4

5.3 Pure Consonants

A pure consonant (i.e., without an implicit vowel) is represented by a two character sequence *cons* 0DCA (*cons* + ්) where *cons* represents a consonant.
e.g. ක් = 0D9A 0DCA

5.4 Consonants with vowel signs

A consonant with a vowel sign is represented by a two character sequence *cons* + *vs* where *vs* represents a vowel sign.
e.g. කා = 0D9A 0DCF, කෞ = 0D9A 0DDE, කෛ = 0D9A 0DDB

Although the al-lakuna and the paa-pillas take two forms, depending on the associated consonant, both forms are represented by the same code character.
e.g. ක් = 0D9A 0DCA, ච් = 0DA7 0DCA
ඌ = 0DB1 0DD4, කූ = 0D9A 0DD4, ඌ = 0DB1 0DD6, කූ = 0D9A 0DD6

	0D8x	0D9x	0DAx	0DBx	0DCx	0DDx	0DEx	0DFx
0		ප෧ 0D90	ච 0DA0	ධ 0DB0	ව 0DC0	ෘ 0DD0		
1		ඵ 0D91	ජ 0DA1	න 0DB1	ශ 0DC1	ෟ 0DD1		
2	ං 0D82	බ 0D92	ඣ 0DA2		ෂ 0DC2	ෲ 0DD2		aa 0DF2
3	ඃ 0D83	භ 0D93	ඤ 0DA3	ඳ 0DB3	ස 0DC3	ෳ 0DD3		ෳ 0DF3
4		ම 0D94	ඥ 0DA4	ප 0DB4	හ 0DC4	෴ 0DD4		෴ 0DF4
5	අ 0D85	ඹ 0D95	ඦ 0DA5	ඵ 0DB5	ළ 0DC5			
6	ආ 0D86	ය෧ 0D96	ජ 0DA6	බ 0DB6	ෆ 0DC6	෵ 0DD6		
7	ඇ 0D87		ට 0DA7	භ 0DB7				
8	ඈ 0D88		ඨ 0DA8	ම 0DB8		a 0DD8		
9	ඉ 0D89		ඩ 0DA9	ඹ 0DB9		ෙ 0DD9		
A	ඊ 0D8A	ක 0D9A	ඪ 0DAA	ය 0DBA	් 0DCA	ේ 0DDA		
B	උ 0D8B	ඛ 0D9B	ණ 0DAB	ර 0DBB		ෛ 0DDB		
C	ඌ 0D8C	ග 0D9C	ඬ 0DAC			ො 0DDC		
D	ඍ 0D8D	ඝ 0D9D	ත 0DAD	ල 0DBD		ෝ 0DDD		
E	ඎ 0D8E	ඞ 0D9E	ථ 0DAE			ෞ 0DDE		
F	ඏ 0D8F	ඟ 0D9F	ද 0DAF		ෟ 0DCF	ෟ 0DDF		

FIGURE 1 – The Sinhala Character Encoding

TABLE 4 – Names of the Characters

Code (1)	Character (2)	Name (3)
Various signs		
0D82	ං	Sinhala sign anusvaraya
0D83	ඃ	Sinhala sign visargaya
Independent vowels		
0D85	අ	Sinhala letter ayanna
0D86	ආ	Sinhala letter aayanna
0D87	ඇ	Sinhala letter aeyanna
0D88	ඈ	Sinhala letter aeeyanna
0D89	ඉ	Sinhala letter iyanna
0D8A	ඊ	Sinhala letter iiyanna
0D8B	උ	Sinhala letter uyanna
0D8C	ඌ	Sinhala letter uuyanna
0D8D	ඍ	Sinhala letter iruyanna
0D8E	ඎ	Sinhala letter iruuyanna
0D8F	ඏ	Sinhala letter iluyanna
0D90	ඐ	Sinhala letter iluuyanna
0D91	එ	Sinhala letter eyanna
0D92	ඒ	Sinhala letter eeyanna
0D93	ඓ	Sinhala letter aiyanna
0D94	ඔ	Sinhala letter oyanna
0D95	ඕ	Sinhala letter ooyanna
0D96	ඖ	Sinhala letter auyanna
Consonants		
0D9A	ක	Sinhala letter alpapraana kayanna
0D9B	ඛ	Sinhala letter mahaapraana kayanna
0D9C	ග	Sinhala letter alpapapraana gayanna
0D9D	ඝ	Sinhala letter mahaapraana gayanna
0D9E	ඞ	Sinhala letter kantaja naasikyaya
0D9F	ඟ	Sinhala letter sanyaka gayanna
0DA0	ච	Sinhala letter alpapraana cayanna
0DA1	ඡ	Sinhala letter mahaapraana cayanna
0DA2	ජ	Sinhala letter alpapraana jayanna
0DA3	ඣ	Sinhala letter mahaapraana jayanna
0DA4	ඤ	Sinhala letter taaluja naasikyaya
0DA5	ඥ	Sinhala letter taaluja sanyooga naaksikyaya
0DA6	ඦ	Sinhala letter sanyaka jayanna
0DA7	ට	Sinhala letter alpapraana ttayanna
0DA8	ඨ	Sinhala letter mahaapraana ttayanna
0D9A	ඩ	Sinhala letter alpapraana ddayanna
0DAA	ඪ	Sinhala letter mahaapraana ddayanna
0DAB	ණ	Sinhala letter muurdhaja nayanna
0DAC	ඬ	Sinhala letter sanyaka ddyanna
0DAD	ත	Sinhala letter alpapraana tayanna
0DAE	ථ	Sinhala letter mahaapraana tayanna
0DAF	ද	Sinhala letter alpapraana dayanna

TABLE 4 – Names of the Characters (Concluded)

Code (1)	Character (2)	Name (3)
0DB0	ධ	Sinhala letter mahaapraana dayanna
0DB1	න	Sinhala letter dantaja nayanna
0DB3	ඳ	Sinhala letter sanyaka dayanna
0DB4	ප	Sinhala letter alpapraana payanna
0DB5	ඵ	Sinhala letter mahaapraana payanna
0DB6	බ	Sinhala letter alpapraana bayanna
0DB7	භ	Sinhala letter mahaapraana bayanna
0DB8	ම	Sinhala letter mayanna
0DB9	ඹ	Sinhala letter amba bayanna
0DBA	ය	Sinhala letter yayanna
0DBB	ර	Sinhaya letter rayanna
0DBD	ල	Sinhala letter dantaja layanna
0DC0	ව	Sinhala letter vayanna
0DC1	ශ	Sinhala letter taaluja sayanna
0DC2	ෂ	Sinhala letter muurdhaja sayanna
0DC3	ස	Sinhala letter dantaja sayanna
0DC4	හ	Sinhala letter hayanna
0DC5	ළ	Sinhala letter muurdhaja layanna
0DC6	ෆ	Sinhala letter fayanna
Sign		
0DCA	්	Sinhala sign al-lakuna
Dependent vowel signs		
0DCF	ා	Sinhala vowel sign aela-pilla
0DD0	ැ	Sinhala vowel sign ketti aeda-pilla
0DD1	ෑ	Sinhala vowel sign diga aeda-pilla
0DD2	ි	Sinhala vowel sign ketti is-pilla
0DD3	ී	Sinhala vowel sign diga is-pilla
0DD4	ු	Sinhala vowel sign ketti paa-pilla
0DD6	ූ	Sinhala vowel sign diga paa-pilla
0DD8	a	Sinhala vowel sign gaetta-pilla
0DD9	ෙ	Sinhala vowel sign kombuva
0DDA	ේ	Sinhala vowel sign diga kombuva
0DDB	ෛ	Sinhala vowel sign kombu deka
Two-part dependent vowel signs		
0DDC	ොා	Sinhala vowel sign kombuva haa aela-pilla
0DDD	ෝ	Sinhala vowel sign kombuva haa diga aela-pilla
0DDE	ෙෘ	Sinhala vowel sign kombuva haa gayanukitta
Dependent vowel sign		
0DDF	ෘ	Sinhala vowel sign gayanukitta
Additional dependent vowel signs		
0DF2	aa	Sinhala vowel sign diga gaetta-pilla
0DF3	ෟ	Sinhala vowel sign diga gayanukitta
Punctuations		
0DF4	෴	Sinhala punctuation kunddaliya

5.5 Non-standard letters

The following letters are represented as shown:
ඍ = 0DBB 0DD0, ඎ = 0DBB 0DD1
ඏ = 0DBB 0DD4, ඐ = 0DDB 0DD6
ඓ = 0DC5 0DD4, ඔ = 0DC5 0DD6

NOTES:

1. These are only the internal representations and not the keyboard sequences.

2. The representation of a letter such as කෝ by a consonant character (ක) followed by several vowel sign characters (ෙ + ා + ්) is permitted in Unicode, but is discouraged in this standard. This standard recommends the use of a single composite vowel sign character following a consonant in such cases (e.g. ක + ෝ) .

3. This standard does not specify a method of displaying the vocalic strokes of certain letters such as කු, ී and ඇ without their associated consonant symbols.

5.6 Semi-consonant signs

An *anusvaraya* (ං) or *visargaya* (ඃ) sign may follow a vowel, a consonant or a vowel sign.
e.g. අං = 0D85 0D82 (අ + ං), අඃ = 0D85 0D83 (අ + ඃ)
කං = 0D9A 0D82 (ක + ං), කෝං = 0D9A 0DDD 0D82 (ක + ෝ + ං),
කූඃ = 0D9A 0DD4 0D83 (ක + ූ + ඃ)

NOTE : *A semi-consonant, if present, is always the last character in a combining character sequence.*

5.7 Rakaaraansaya and Yansaya

The rakaaraansaya and yansaya are forms of **conjunct letters**.

The rakaaraansaya ් represents a ර which follows a pure consonant. It can, in turn, be followed by a vowel sign. It is joined to the preceding letter by a zero-width joiner (zwj).

A rakaaraansaya is represented by the character sequence *cons* 0DCA 200D 0DBB
(*cons* + ් + zwj + ර) where *cons* represents some consonant.
e.g.: ක්‍ර = 0D9A 0DCA 200D 0DBB (ක + ් + zwj + ර),
ක්‍රෙ = 0D9A 0DCA 200D 0DBB 0DD9 (ක + ් + zwj + ර + ෙ)

Similarly the yansaya ්‍ය represents a ය which follows a pure consonant.
e.g. ක්‍ය = 0D9A 0DCA 200D 0DBA (ක + ් + zwj + ය)
කොය් = 0D9A 0DCA 200D 0DBA 0DDD (ක + ් + zwj + ය + ෝ)

NOTES:

1. *As the ් and ්‍ය are present on the keyboard, users will not need to key in the above sequences (see **6**).*

2. *The yansaya and rakaaraansaya are required in normal Sinhala text. However, if for some reason, it is desired not to use the rakaaraansaya or yansaya, the zwj should be omitted.*

5.8 Repaya

The repaya ් represents the letter ර preceding a consonant. It is represented by the sequence 0DBB 0DCA 200D *cons* (ර + ් + zwj + *cons*).
e.g. කර්‍ම = 0D9A 0DBB 0DCA 200D 0DB8 (ක + ර + ් + zwj + ම)

A ya followed by a yansaya with a repaya (in words such as කාර්‍යාලය) is represented by the code sequence 0DBB 0DCA 200D 0DBA 0DCA 200D 0DBA (ර + ් + zwj + ය + ් + zwj + ය).

NOTE : *As the repaya appears on the keyboard, users will not need to key in the above sequences.*

5.9 Other conjunct letters (බැඳි අකුරු)

Conjuct letters are represented by the sequence *cons 0DCA 200D cons* (*cons* + ් + zwj + *cons*). The second consonant may optionally be followed by a vowel sign.
e.g. න්‍ද = 0DB1 0DCA 200D 0DAF (න + ් + zwj + ද),
කෙෂ්‍ඨ = 0D9A 0DCA 200D 0DC2 0DDA (ක + ් + zwj + ෂ + ෙ ්)
Conjuct letters may be further joined by a rakaaraansaya or yansaya.
e.g. ත්‍ර්‍ය = 0DB1 0DCA 200D 0DAF 0DCA 200D 0DBB 0DCF
(න + ් + zwj + ද + ් + zwj + ර + ්)

5.10 Touching letters

A pure consonant written touching the following letter instead of using an al-lakuna, used in classical and Buddhist texts, is represented by the sequence *cons* 200D 0DCA *cons*.
e.g. ඤ්‍ඤ= 0DC3 200D 0DCA 0DC3 (ඤ + zwj + ් + ඤ)

5.11 Stand-alone signs

A vowel sign without an associated consonant may be displayed by preceding it with an INV character. e.g. ා = INV 0DCF (INV + ා), ෙ= INV 0DD9 (INV + ෙ). The INV character should not have any width when combining with a pre-base or .post-base sign, and should have a default width when combining with an above- or below-base sign.

A stand-alone yansaya is represented by the sequence INV 0DCA 200D 0DBA (INV + ් + zwj + ය). A stand-alone rakaaraansaya is represented similarly.

A stand-alone repaya is represented by the sequence 0DBB 0DCA 200D INV (ර + ් + zwj + INV).

The code for the INV character is not defined in this standard, as it is still under consideration by Unicode. Until such time such a code is inlcluded in Unicode, implementers may use the non-break space (NBSP) character (code 00A0) as the INV character, keeping in mind that this may change.

6 KEYBOARD INPUT

Text encoded as specified in this standard may be input to a computer system in many ways, e.g., text recognition, 10-key keypad, etc. However, much text is input using standard computer keyboards (e.g., 101-key). This section provides guidelines on how Sinhala text may be entered using such a keyboard.

Each Sinhala letter, such as ඖ (represented by a character sequence) is input by a sequence of keys on the keyboard. Thus there is a many-to-many relation between keys and characters.

The set of symbols which appear on the keyboard, and the key sequences to generate each letter, are independent of the mapping of symbols to keys. The same key sequences may be used by several keyboard layouts.

This section specifies the recommended key sequences for generating Sinhala characters on a standard computer keyboard. The recommended keyboard layout, based on the Wijesekara keyboard, is specified in **7**.

Key sequences are defined on the principle “type as you write”. Each symbol is typed in the order it is written in, which may be different from the encoding sequence or the display order.

NOTE : *This standard does not specify the symbols which are displayed during the intermediate stages in the construction and deletion of letters.*

6.1 Keyboard Symbols

A Sinhala keyboard should have the following symbols, which may be assigned to keys in suitable ways. Each physical key may have several symbols assigned to it, one for each shift-state. Generally, only the symbols of the unshifted and the normal shift states will be printed on the keyboard.

Keys should be assigned to the following:

a) Consonants

ක ඛ ග ඝ ඞ ඟ
ච ඡ ජ ඣ ඤ ඦ ඥ
ට ඨ ඩ ඪ ණ ඬ
ත ථ ද ධ න ඳ
ප ඵ බ භ ම ඹ
ය ර ල ව ශ ෂ
ස හ ළ ෆ

b) Vowels

අ ඉ ඊ උ ඍ ඏ එ and ඔ

NOTES :

1. Other vowels are produced by a key sequence.

2. The ඏ symbol need not be printed on the keyboard, but may be keyed using a shift state.

c) Vocalic Strokes

Vocalic strokes comprise the leading vowel sign: kombuwa (ෙ), and the trailing vowel signs: aela-pilla (ා), ketti aeda-pilla (ැ), diga aeda-pilla (ෑ), ketti is-pilla (ි), diga is-pilla (ී), keti paa-pilla (ු), diga paa-pilla (ූ), gaeta-pilla (ෘ) and gayanukitta (්).

NOTE : *Composite vowel signs are entered as a sequence of two or more keys.*

d) Other Symbols

The al-lakuna**:** (්), semi-consonants: anusvaraya (ං) and visargaya (ඃ), non-vocalic strokes: yansaya (්‍ය), rakaaraansaya (්‍ර) and repaya (්‍ර), punctuation: kundaliya (෴), the letter muurdhaja lu (ළු), the non-breaking space: used to enter intra-word space e.g., in ගියේ ය, and keys for non-Sinhala symbols such as numerals, punctuation marks and standard symbols.

NOTES:

1. *Although the **al-lakuna** takes two forms, (e.g. ක් and ච්) they are both entered using the same key.*

2. *As the letter ළු has a non-standard form, it is assigned a key for user convenience.*

6.2 Special Keys

The following special keys are specified. The sanyakaya may be used to generate “sanyaka” letters such as ඟ and ඳ in conjunction with letters such as ග and ද. The join key is used to join two letters to form conjunct letters such as ඤ්ඣ. The touch key is used to bring two letters together to form “touching” letters, e.g., ඬ The inv key is used to produce an invisible base character.

NOTES:

1. *Keys are also directly assigned to the symbols ඟ, ඳ etc.*

2. *An implementer may assign a key for the zero-width joiner character, although it is not required to enter Sinhala text.*

6.3 Key sequences

Each letter is entered by one or more key sequences as follows:

a) Consonants

A consonant is entered with a single key.
e.g. ක ම ඬ, ඟ

b) Vowels

A vowel is entered with 1 or 2 keys.

අ
අ + ා = ආ
අ + ැ = ඇ
අ + ෑ = ඈ
ඉ
ඊ
උ
උ + ූ = ඌ
සිa
සිa + a = සිaa
එ
එ + ් = ඒ
ෙ + එ = ඓ
ඔ
ඔ + ් = ඕ

ඔ + ෟ = ඕෟ

c) Pure consonants

A pure consonant is entered with 2 keys: *cons* + al-lakuna.
e.g. ක + ් = ක්, ම + ් = ම්

NOTE : *The same key is used for both types of al-lakuna*.

d) Vowel signs

The aela-pilla is entered with 1 key: *cons* + aelapilla.
e.g. ක + ා = කා, ද + ා = දා

The aeda-pili are entered with 1 key: *cons* + ketti aeda-pilla, *cons* + diga aedapilla.
e.g. ක + ැ = කැ, ක + ෑ = කෑ

The is-pili are entered with 1 key: *cons* + keti ispilla, *cons* + diga ispilla.
e.g. ක + ි = කි, ක + ී = කී

The paa-pili are entered with 1 key: *cons* + diga paapilla, *cons* + keti paapilla.
e.g. ප +ු = පු, ප +ූ = පූ

The gaetta-pili are entered with 1 or 2 keys: *cons* + gaetapilla, *cons* + gaetapilla + gaetapilla.
e.g. ග + ෘ = ගෘ, ග + ෘ + ෘ = ගෲ

The kombuva is entered with 1 key, before the consonant: kombuwa + *cons.*
e.g. ෙ + ක = කෙ

The diga-kombuva is entered with 2 keys, before and after the consonant: kombuwa + *cons* + al-lakuna. e.g. ෙ + ක + ් = කේ, ෙ + ම + ් = මේ

The kombu-deka is entered with 2 keys, before the consonant: kombuwa + kombuwa + *cons*. e.g. ෙ + ෙ + ක = ෙකෙ

The kombuva haa aela-pilla is entered with 2 keys before and after the consonant: kombuwa + *cons* + aelapilla. e.g. ෙ + ක + ා = කො

The kombuva haa diga aela-pilla is entered with 3 keys before and after the consonant: kombuwa + *cons* + aelapilla + al-lakuna. e.g. ෙ + ක + ා + ් = කෝ

The kombuva haa gayanukitta is entered with 2 keys before and after the consonant: kombuwa + *cons* + gayanukitta. e.g. ෙ + ක + ෟ = කෞ

NOTES:

1. *There are no special keys for ර and ර්, e.g.: ර + ැ = ර, ර + ෑ = ර්*
2. *The kombuwa key is pressed before the consonant, in writing order.*
3. *The paa-pili for ක ඛ ග, etc. are entered using the same keys as for the other letters. e.g.: ක +ු = කු*
4. *The character ළු is assigned to a key. However it can also be entered as ළ +ු = ළු.*
5. *The character ළූ may be keyed as either ළ +ූ = ළූ, (logical sequence) or ළු + ෑ = ළූ (visual sequence).*
6. *The characters ර්‍ය and ර්‍ය are entered as ර +් and ර +් respectively.*

7. There is no key for the sign ◌aa, it is entered by pressing the a key twice.

e) Semi-consonants

The signs ◌ and ◌ are keyed following a consonant or vowel sign. They will always be the last key of a key sequence.

6.4 Conjunct letters

a) Rakaaraansaya (◌)

This symbol will normally be entered using the rakaaraansaya key.
e.g.: ක +◌ = ක්‍ර

Vowel signs, as shown in Table 3, may be keyed following the rakaaraansaya. However, the kombuwa is keyed preceding the base letter. The ispilla may be keyed either before or after the rakaaraansaya, to conform with the practice in writing.
e.g. ට + ◌ = ට්‍ර, ට +◌ + ා = ට්‍රා , ට +◌ + ෑ = ට්‍රෑ, ට +◌ + ෑ = ට්‍රෑ,
ත +◌ +ි = ත්‍රි, ෙ + ප +◌ + ා = ෙප්‍රා
The following alternative sequences are also valid:
e.g. බ +ි +◌ = බ්‍රි, ට +ි +◌ = ට්‍රි

NOTE : *Sequences other than those shown in Table 3, such as ට +◌ +ෘ and ට +◌ +ෟ are not used with the rakaaraansaya, but keying them in should be allowed.*

b) Yansaya (◌)

This symbol will normally be entered using the yansaya key.
e.g. ත + ◌ = ත්‍ය
The yansaya may be combined with vowel signs, as shown in Table **3**.
e.g. ක + ◌ +ෟ = ක්‍යූ, ෙ + ක + ◌ = ෙක්‍ය
Sequences other than those shown in Table **3** are not used with the yansaya, but keying them in should be allowed.

c) Repaya

The repaya symbol is keyed in following a consonant, i.e., the writing sequence.
e.g. ක + ◌ = ර්‍ක

d) Other conjunct letters (බැඳි අකුරු)

Conjunct letters are generated by pressing the first consonant, the "join" key and the 2nd consonant. The al-lakuna is not typed.

e.g.: ක + join + ෂ = ක්ෂ, ක + join + ෂ + ා = ක්ෂා
ෙ + ක + join + ෂ + ා + ◌ = ෙක්ෂෝ

6.5 Other

a) Sanyaka letters

The sanyaka letters ඳ ඟ ඬ ඣ may be generated by the keys ද ග ඩ ජ respectively, followed by the sanyakaya key. This is an optional feature, as the sanyaka letters may also be entered directly.

b) Touching letters

Touching letters are entered by pressing the first consonant, the “touch” key, and the 2nd consonant, e.g., ස + touch + ස = සස.

c) Stand-alone signs

Stand-alone vowel signs are keyed using the "inv" key (○) followed or preceded by the desired vowel signs or other symbol.

The sequence ya yansaya repaya (as in ආය්‍ය) is entered by the key sequence ය + ්‍ය + ්‍ය.

7 KEYBOARD LAYOUT

The keyboard symbols specified in **6** are mapped to a keyboard by the keyboard layout given in Figure **2**. It is a modification of the layout in **SLS 1134:2001** which in turn was based on the Wijesekara keyboard.

7.1 Keys

A standard computer keyboard has 48 assignable keys on the main keyboard. Each physical key on a standard computer keyboard can be assigned to up to 4 symbols, for the following shift states:

unshifted
shift
ctrl-alt (alt-gr)
shift ctrl-alt (shift alt-gr)

Therefore a total of 48 × 4 = 196 keys are available for use. The symbols defined in **6** are assigned to the unshifted and shifted states of the keys as shown in Figure **2**.

NOTE : *The shift-ctrl-alt state is not used in this keyboard layout, but implementers may use it for entering additional symbols, etc.*

7.2 Keyboard Layout

The layout of the keyboard is shown in Figure **2.**

Table 5 – Special keys and Symbols

Symbols Used	**Additional keys** (not shown on keyboard)	**Infrequently Used Symbols**
්‍ය = yansaya ◌ෟ = rakaaraansaya ◌් = repaya ◌‍◌ = join adjacent letters The shifted form of this key produces "touching" letters	ඳ = alt-gr-ද (alt-gr-o) ඟ = alt-gr-ග (alt-gr-.) ඬ = alt-gr-ඩ (alt-gr-v) ඹ = alt-gr-බ (alt-gr-c) ඃ = alt-gr- ං (alt-gr-x) non-breaking space = shift-space ○ (invisible) = alt-gr-space	ඵ = alt-gr-ඵ ◌ඁ = alt-gr-◌ (alt-gr-a) ◌ං = alt-gr-. (alt-gr-') **Optional** sanyakaya = alt-gr-z

NOTE : *The marks for strokes such as ◌ෘ , ◌ෲ , ◌ෟ and ◌ෳ in the keyboard layout also stand for other equivalent strokes whose shapes vary according to the associated consonant. Please see **5.4** for details.*

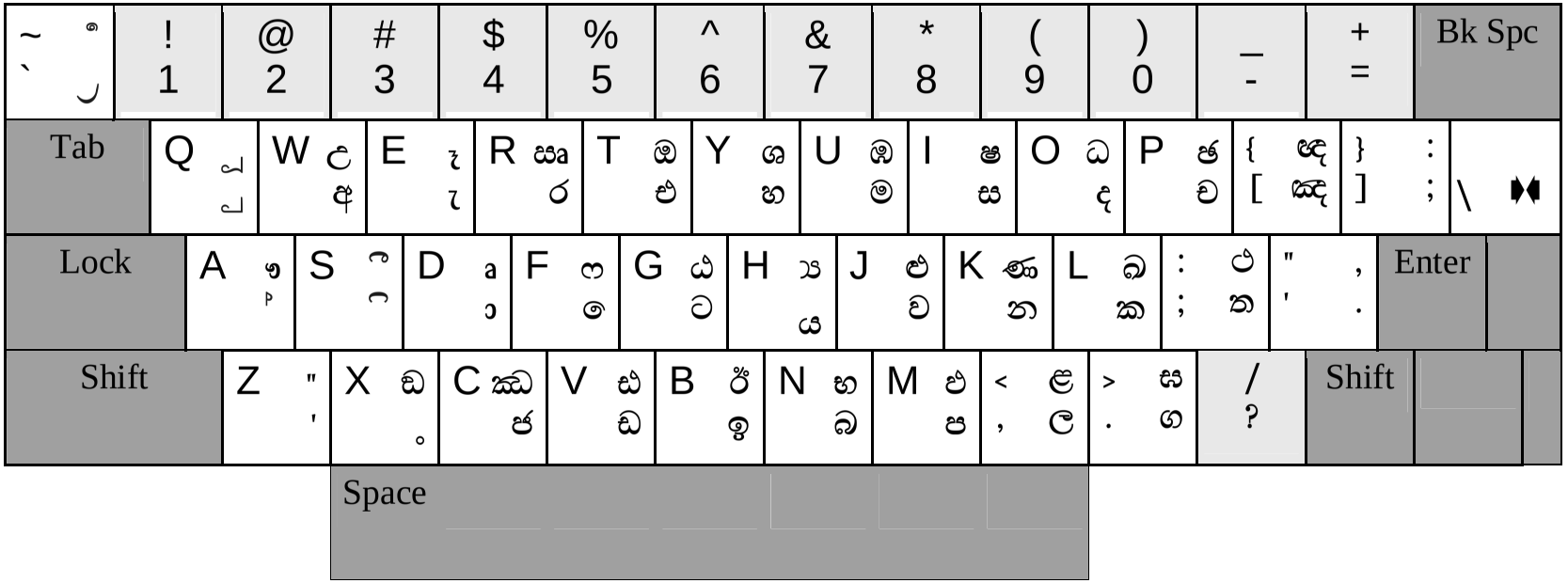

FIGURE 2 – Recommended Standard Keyboard Layout

SLS CERTIFICATION MARK

The Sri Lanka Standards Institution is the owner of the registered certification mark shown below. Beneath the mark, the number of the Sri Lanka Standard relevant to the product is indicated. This mark may be used only by those who have obtained permits under the SLS certification marks scheme. The presence of this mark on or in relation to a product conveys the assurance that they have been produced to comply with the requirements of the relevant Sri Lanka Standard under a well designed system of quality control inspection and testing operated by the manufacturer and supervised by the SLSI *which includes surveillance inspection of the factory, testing of both factory and market samples.*

Further particulars of the terms and conditions of the permit may be obtained from the Sri *Lanka Standards Institution, 17, Victoria Place, Elvitigala Mawatha, Colombo 08.*

Printed at SLSI (Printing Unit)

SRI LANKA STANDARDS INSTITUTION

The Sri Lanka Standards Institution (SLSI) is the National Standards Organization of Sri Lanka established under the Sri Lanka Standards Institution Act No. 6 of 1984 which repealed and replaced the Bureau of Ceylon Standards Act No. 38 of 1964. The Institution functions under the Ministry of Science & Technology.

The principal objects of the Institution as set out in the Act are to prepare standards and promote their adoption, to provide facilities for examination and testing of products, to operate a Certification Marks Scheme, to certify the quality of products meant for local consumption or exports and to promote standardization and quality control by educational, consultancy and research activity.

The Institution is financed by Government grants, and by the income from the sale of its publications and other services offered for Industry and Business Sector. Financial and administrative control is vested in a Council appointed in accordance with the provisions of the Act.

The development and formulation of National Standards is carried out by Technical Experts and representatives of other interest groups, assisted by the permanent officers of the Institution. These Technical Committees are appointed under the purview of the Sectoral Committees which in turn are appointed by the Council. The Sectoral Committees give the final Technical approval for the Draft National Standards prior to the approval by the Council of the SLSI.

All members of the Technical and Sectoral Committees render their services in an honorary capacity. In this process the Institution endeavours to ensure adequate representation of all view points.

In the International field the Institution represents Sri Lanka in the International Organization for Standardization (ISO), and participates in such fields of standardization as are of special interest to Sri Lanka.

Printed at the Sri Lanka Standards Institution, 17, Victoria Place, Elvitigala Mawatha, Colombo 08.