

Writing Sinhala Using an English Keyboard

S. C. Jayewardena

1. Introduction

In this age of information, it is vital that we can make available the information we have to other people. Similarly, we must be able to access the information the other people have. This is most efficiently done using computers which can be used to store large amounts of data in a relatively small area. People all around the world use ASCII characters to store all kinds of textual data, usually as English text. Using computer networks, thousands of people access such documents.

Many people will profit if Sinhala textual data could also be placed on such a network. These people will include not only academics but people of all professions. Sinhala speaking people outside Sri Lanka could also access these data.

This clearly requires some convention as to how a particular Sinhala word is written using the English alphabet. This paper describes a particular scheme called *sumihiri* designed for this purpose.

2. Definition of Atomic Letters

We first define what we call *atomic letters* of *sumihiri*. These atomic letters stand for the usual "a" letters (consonants) of Sinhala. Here is the complete list:

ක	k	ඌ	th	ශ	sh
ක	K	ඌ	Th	ඡ	Sh
ග	g	ඌ	dh	ජ	s
ග	G	ඌ	Dh	ඣ	h
ච	Z	ඌ	n	ඤ	l
ච	ch	ඌ	p	ඦ	Ndh
ඡ	Ch	ඌ	P	ට	Nd
ඣ	j	ඌ	b	ඨ	Ng
ඣ	J	ඌ	B	ඩ	Mb
ඤ	q	ඌ	m	ඪ	f
ඤ	t	ඌ	y	ණ	GN
ඦ	T	ඌ	r		x
ඦ	d	ඌ	L		
ට	D	ඌ	w		
ඨ	N				

3. Vowel Letters

Then we have what are called *vowel letters* of *sumihiri*. These are shown below:

අ	a	ඌ	uu
ආ	aa	ඍ	E
ඈ	z	ඎ	ee
ඉ	zz	ඏ	o
ඊ	i	ඐ	oo
උ	ii	එ	ai
ඌ	u	ඒ	au

Most of them are what you expect intuitively form the sound of the letter. There is no letter in English that uniquely corresponds to the sound of "අ". The letter "z" has been chosen, because it at least looks like "ඈ" which is used in Sinhala to get that sound. The natural candidate for the sound of "ඍ" would be "e". However, *sumihiri* reserves it for some other purpose that will be explained shortly.

The extended sounds are obtained by repeating the basic letter. For instance, "a" stands for "අ" and "aa" stands for "ආ". The only exception is "ඵ" which is obtained by repeating "e" but not "E" twice. Typing "ee" is easier than typing "EE"; that is the reason.

4. Constructing Base Letters

Base letters are obtained by combining atomic letters with vowel letters. The following table shows some examples:

	a/e	aa	z	zz	i	ii	u	uu	E	ee	o	oo	ai	au
k	ක	කා	කෑ	කෑ	කී	කී	කු	කු	කෙ	කේ	කො	කෝ	කෙ	කො
g	ග	ගා	ගෑ	ගෑ	ගී	ගී	ගු	ගු	ගෙ	ගේ	ගො	ගෝ	ගෙ	ගො
d	ධ	ධා	ධෑ	ධෑ	ධී	ධී	ධු	ධු	ධෙ	ධේ	ධො	ධෝ	ධෙ	ධො
dh	ඳ	ඳා	ඳෑ	ඳෑ	ඳී	ඳී	ඳු	ඳු	ඳෙ	ඳේ	ඳො	ඳෝ	ඳෙ	ඳො
Dh	ධ	ධා	ධෑ	ධෑ	ධී	ධී	ධු	ධු	ධෙ	ධේ	ධො	ධෝ	ධෙ	ධො
r	ර	රා	රෑ	රෑ	රී	රී	රු	රු	රෙ	රේ	රො	රෝ	රෙ	රො
L	ළ	ලා	ළෑ	ළෑ	ළී	ළී	ලු	ලු	ළෙ	ළේ	ළො	ළෝ	ළෙ	ළො
Ndh	ඳ	ඳා	ඳෑ	ඳෑ	ඳී	ඳී	ඳු	ඳු	ඳෙ	ඳේ	ඳො	ඳෝ	ඳෙ	ඳො

The table above indicates that adding either "a" or "e" both result in the same letter. That is, "ka" and "ke" both result in "ක". Although they represent the same Sinhala letter, in *sumihiri* they are pronounced differently. The first one is an *open* "ක" whereas the second one is a *closed* "ක". A good example would be the two "ම" letters in the word "මම". The distinction between the two sounds makes reading *sumihiri* much easier. This is the reason why "e" was not used to produce the "ඵ" sound.

5. "අ:" and "ආ:"

There are two special letters for this:

අ:	NG	ආ:	H
----	----	----	---

6. "-R", "-Y" and "R-" Letters

Atomic letters can be *extended* in two ways. They can be appended by "R" or "Y". They can also be prepended by "R". Once extended this way, they can again be combined with the vowel letters to produce the Sinhala letters not described so far. The following tables show what can be obtained by the extended atomic letter "k".

ඉ	kR	කෙ	kRE
ඉ	kRa/ke	කේ	kRee
ඉා	kRaa	කො	kRo
ඉැ	kRz	කෝ	kRoo
ඉෑ	kRzz	කො	kRai
ඉි	k R i	කො	kRau
ඉී	kRii		
කා	k R u		
කෘ	k R u u		

කඞ	kY	කො	kYE
කඞ	kYa/kYe	කේ	kYee
කඞා	kYaa	කො	kYo
කඞැ	kYz	කෝ	kYoo
කඞෑ	kYzz	කො	kYai
කඞි	k Y i	කො	kYau
කඞී	k Y i i		
කඞු	k Y u		
කඞූ	k Y u u		

ඊ	Rk	කේ	RkE
ඊ	Rka/Rke	කේ	Rkee
ඊා	Rkaa	කො	Rko
ඊැ	Rkz	කෝ	Rkoo
ඊෑ	Rkzz	කො	Rkai
ඊි	R k i	කො	Rkau
ඊී	R k i i		
ඊු	R k u		
ඊූ	R k u u		

Note that, although an extended atomic letter may be combined with any of the vowel letters, that does not always result in a meaningful Sinhala letter. Look, for example, at the combination "kYau" above. It is upto the writer of the text, to write meaningful letters.

7. Problems

Although *sumihiri* is extremely versatile, there are certain rare cases, in which you need some way to escape *sumihiri*'s normal interpretation.. Suppose you want to write (for whatever reason) the sequence of letters "අඉ". Now if you simply write "ai" this will be interpreted as "ඵ", which is not what you wanted.

A more serious situation occurs in conjunction with nasal letters that start with "N"; that is with "Nge", "Nde" and "Ndhe". For example, suppose you want to write "මණ්ඩලය". According to the above rules you write "maNdeleya". However *sumihiri* will interpret this as "මඩලය" !

The solution to this kind of misinterpretation is to insert an underscore ("_") to show the letter boundary.

Example:

මණ්ඩලය අඉ	maN_deleye a-i
--------------	-------------------

Theoretically, this kind of problem can occur with any of the following multi-character atomic letters: th, Th, dh, Dh, sh, Ndh, Nd, Ng, GN. However, a little thought will convince you that, practically, the problem occurs only with "Nd", which is also a rare case.

8. Discussion

One big plus point for *sumihiri* is that there is a great uniformity in it. Consider, for example, the letters "ඃ", "ක", "ර", and "ඵ". Note how the first three of them are obtained by adding

different decorations to the base letters "භ", "ක" and "උ". In the case of "ඵ", it cannot even be obtained by adding a decoration to "භ" or "භ". But in *sumihiri* they are written in a very uniform manner; namely as "su", "ku", "ru" and "Lu". Further examples:

pz භ	rz උ	s ඵ	m ම
------	------	-----	-----

Another important fact is that, unlike many other schemes that have been proposed to code Sinhala using English alphabet, *sumihiri* does not attempt to further reduce the richness of the written Sinhala language, just for the sake of the scheme. The scheme *sumihiri* accomodates everything the present-day written Sinhala requires.-

Those people who can understand Sinhala, but cannot read (Sinhala) can profit form *sumihiri*, because they will easily be able to reproduce the sounds of the Sinhala words without having to know the real Sinhala alphabet. Similarly, *sumihiri* would be ideal for the foreigner learning the Sinhala pronunciation and writing.

It is even conceivable that a computer (robot) can easily be taught to read aloud a Sinhala text written in *sumihiri*.

Finally, *sumihiri* can actually be considered as a complete replacement for the written Sinhala, although it was not the author's intention in designing *sumihiri*. One simplification that may be used when writing *sumihiri* by hand, or even with a typewriter would be to use a *bar* over the letters "a", "z", "i", "u", "e", "o" when they are repeated to get the extended sound.