

Romanized Myanmar Input Method for Mobile Devices

Hlaing Myint Oot†

Katsuko T. Nakahira†

Yoshiki Mikami†

†Nagaoka University of Technology

Introduction

Input methods for Myanmar Language in electronic devices, like computer and mobile phone, are not standardized and the user has to memorize the character positions. This hinders widespread use of Myanmar-based SMS. Here, we introduce Romanized Myanmar language input method in which user, even non-native speaker, can input pronunciation of words with Roman script, instead of typing original Myanmar spelling of words. This method is an interactive system and similar to Japanese Kana-Kanji conversion, through which user chooses the correct one from the list of candidates. In order to create the efficient database with less memory, we made survey on native speakers.

Background and Objective

Recently, there is no standard rule for writing Myanmar word in roman script. Many scholars have tried to make one but none of them have been successful yet. For Romanization of Myanmar language, it is very difficult to show accurate pronunciation of a word. And there are many differently spelling by users of different sex, age, region, educational level and English proficiency. For example, a word “ဇေ” [phay] which is a part of person name, is found in various spellings such as:- Hpei, Pe, Phae, Phay, Phei, etc. This can also be found in writing many other names, syllables, and words.

In order to make Myanmar word readable for non-native speaker, transcription method is used to write the word in phonetic form. This method is generally used in a dictionary by implementing an international standard such as the IPA. A word, for example, “မိုးမြင့်” is written as “mi: gji”. But it is not widely used in daily life as many special symbols are introduced and the knowledge of phonetics is needed.

The objective is to develop Myanmar Language Input system in Romanization Method using only Roman characters without any special symbols implementing on mobile phone which is diffusing much faster than personal computers in Myanmar.

Concept of Input System on mobile phone

The concept of this input method is to input pronunciation of Myanmar syllables, that user is thinking, in Roman script [1]. And user is able to select right combination from candidates which is produced by the system, appear on the selection menu as shown in the Fig.1.

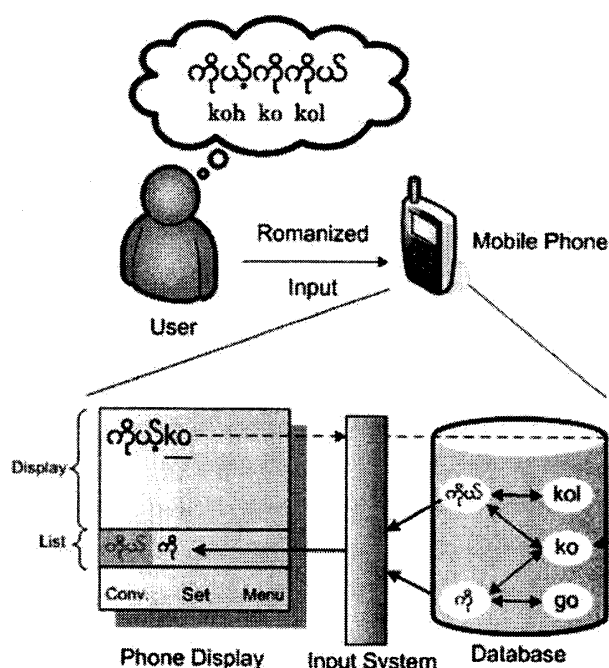


Fig.1 Concept of Romanized Myanmar Input

In order to decide the priority of the syllable candidates, we have to find the probability of each syllable. The probability of Roman spelling appearing as Myanmar syllable can be defined as in Equation (1).

$$P(r_j) = P(m_i) \cdot P(m_i | r_j) \quad (1)$$

where, $r_j = l^*, l \in L$

r_j : Romanized Myanmar spelling

$L = \{a, b, \dots, z\}$

m_i : Myanmar syllable

$P(r_j)$: Probability that a specific Roman spelling r_j appears key-in strokes

$P(m_i)$: Probability of a specific Myanmar syllable m_i appears in a Myanmar text

$P(m_i|r_j)$: Probability of Myanmar syllable m_i is spelled into Roman spelling r_j .

Survey and Results

As the Romanized input of Myanmar language varies by user, a survey to get most frequently used Myanmar syllables as well as $P(m_i)$ is done. The survey is done by using the algorithm of Myanmar syllable breaking [2] from the issues of various fields, e.g. Legal statement, Blogs, Online Magazines and 1380 types of syllables are obtained.

For the Survey of evaluating $P(m_i|r_j)$, top 300 frequently used Myanmar syllables (m_i) are selected. Romanized Myanmar spelling (m_i) for each syllable is collected from 20 native Myanmar people of different areas and different education backgrounds. From this survey; we got many different Romanized Myanmar spelling (r_j) data for 300 syllables. By analyzing these data, we got to know that only 31 syllables has uniform Romanized spelling (r_j) and each syllable has 3.5 ways of spelling in average (total r_j : 1098).

Most frequently used Romanized Myanmar spelling pattern for vowels and consonants are collected by two steps. Firstly, the $P(m_i|r_j)$ table of most frequently used Romanized Myanmar spelling for each syllable is created as part of it is shown in the table.1.

Table.1 Part of the $P(m_i|r_j)$ table

	$P(m_i)$	ae	air	El	ae	ko	so	soe
အဲ	0.038	0.5	0.3	0.1	0.1	0	0	0
ကို	0.017	0	0	0	0	1	0	0
နိုး	0.0007	0	0	0	0	0	0.3	0.7
ဆို	0.008	0	0	0	0	0	0.7	0.2

In second step, $P(r_j)$ for each syllable is calculated by multiplying $P(m_i)$ and $P(m_i|r_j)$. The value of $P(r_j)$ is used to make the table of Roman-Myanmar table as part of it is shown in the table.2. So, the syllables are stored in the database by order of bigger $P(r_j)$ value.

Table.2 Example of Roman-Myanmar $P(r_j)$ table

Roman	Myanmar	$P(m_i r_j) \times P(m_i)$
soe	နိုး	$0.7 \times 0.0007 = 0.00049$
	ဆို	$0.2 \times 0.008 = 0.0016$
so	နိုး	$0.3 \times 0.0007 = 0.00021$
	ဆို	$0.7 \times 0.008 = 0.0056$

Conversion method

Romanized Myanmar spelling is divided into two parts which are the consonant and the part except consonant (later this part is called vowel). As a result, we got the two Romanized Myanmar tables for consonant and vowel. In the table for consonant, there are 24 pairs of Roman and Myanmar characters And in the table for vowel, there are 47 pairs of Roman and Myanmar characters in which Myanmar vowels are matched with each Roman character.

By using the data from the tables, we can create the conversion table (some of the examples are shown in table.3) which is used by the input system to show out the syllable candidates on the selection menu depending on the user input.

Table.3 Part of the Conversion table

C/V	a	aik	Aing
b	ဘ ဝ	ဘိုက် ခိုက်	ဘိုင် ခိုင်
d	ဒ ဓ ည ဝ	ဒိုက် ခိုက် ညိုက် ခိုက်	ဒိုင် ခိုင် ညိုင် ခိုင်
f	ဖ	ဖိုက်	ဖိုင်
g	ဂ ဃ	ဂိုက် ဖိုက်	ဂိုင် ဖိုင်

Conclusion and Future Works

In this paper, we have done questionnaires and analysis in order to create the database structure which is based on a probabilistic method. This is very important to decide the priority of the syllable candidates. But there are some tasks to solve as Myanmar language has many different glyphs when the syllables are combined to become a word.

References

- [1] Hlaing Myint Oo, Yoshiki Mikami and Katsuko T. Nakahira, Design of Input Method for Myanmar Language on Mobile Phone, FIT 2008, pp.473-475.
- [2] Zin Maung Maung and Yoshiki Mikami, A Rule-based Syllable Segmentation of Myanmar Text, IJCNLP-08 Workshop on NLP for Less Privileged Languages, Hyderabad, India, pp.51-58, Jan 11, 2008.