

Design of Input Method for Myanmar Language on Mobile Phone

Hlaing Myint Oo†

Katsuko T. Nakahira†

Yoshiki Mikami†

1. Abstract

Mobile phone has become an indispensable part of people's lives in developing countries. Myanmar started SMS in 2005 and several input methods for the Myanmar Language on a mobile phone have been introduced. However, these input methods bear no commonness and burden of the user to memorize the character arrangement hinders widespread use of those. This paper introduces Romanized Myanmar language input method in which user types pronunciation of words with Roman script, instead of typing original Myanmar spelling of words. This method is an interactive system and similar to Kana-Kanji conversion, through which user chooses the correct one from the list of candidates.

2. Introduction

More and more people are using mobile phones and the world is increasingly becoming unwired due to diffusion of this technology. The cellular technology has been present in Myanmar since early 1990s. The number of subscribers is increasing year by year along with the development of SMS. Especially in rural areas with poor ICT infrastructure and lack of electricity, mobile phone is the best communication tool because it can be recharged even from car battery and requires no wired connection. It is expected that the diffusion of the mobile phones will become faster than that of personal computers.

However, Myanmar language input methods implemented in mobile phone bear no commonness and poses difficulty for the user to memorize the character arrangement. And IT knowledge gap is caused in city and provinces because there is no dedicated curriculum for IT until high school. According to the investigation that UNESCO Institute for Statistics had done in September 2006, literacy rate of Myanmar people (15 years and older) turned out 89.9%. The education system in Myanmar requires all the students to learn English language from the elementary school together with Myanmar language. So, one will understand English very well but he or she may not have IT knowledge at all.

As a result, more efficient input method, which is very easy to be used by user without IT knowledge, is necessary. By using the concept introduced in this paper, user needs no IT knowledge and memorization. This

may give a good chance to get a relief of the delay from IT age and the digital divide problems of Myanmar.

3. Myanmar Language and Script

Myanmar language is an official language in Myanmar. It is spoken by majority of population, and as a second language by ethnic minorities in Myanmar. The language is a tonal and analytic language. It is a member of the Tibeto-Burman languages. It utilizes the Myanmar script, derived from the Mon script. Myanmar script is rounded in shape. It is written in sequence from left to right without regular spacing which is sometimes used to mark phrases.

Myanmar script has various characters: consonants, medials, killers and vowels. The basic consonants in Myanmar can be extended by medials. Syllables or words are formed by consonants combining with vowels. However, some syllables can be formed by just consonants, without any vowel. Other characters in the Myanmar script include special characters like Pali, punctuation marks, stacked characters, tones and abbreviations.

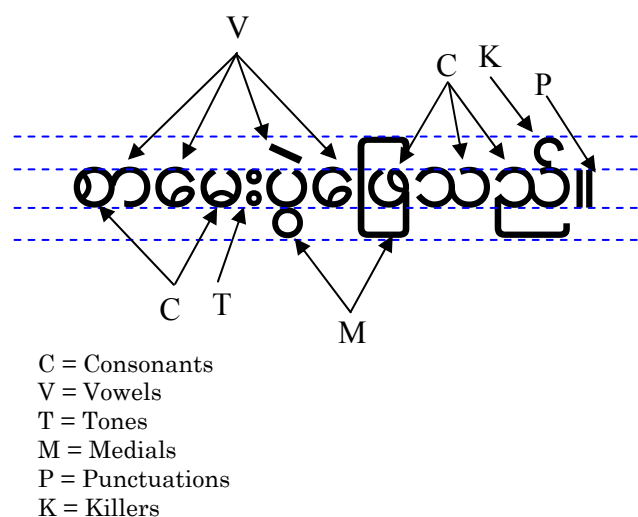


Fig.1 An Example of Myanmar word
“sar mae pyoe pyae thi.”
 (“Sit for the exam” in English)

4. Three Approaches of Input

There are mainly three approaches of Myanmar character input methods on mobile phone's numerical key pad. One of the intrinsic methods is *direct-input*, for example, multi-tapping in which user pushes key

† Nagaoka University of Technology

several times until the desired character appears. The second one is *transcription* method that input the special written symbols, for example, International Phonetic Alphabet, of the sounds of Myanmar Word using a set of exact rules. The last one is called *transliteration*, which creates a mapping from one script to another that is designed to match the original script as directly as possible. In this research, we introduce Romanization method that encompasses both transliteration and transcription methods. The differences among three approaches are shown in the Fig.2.

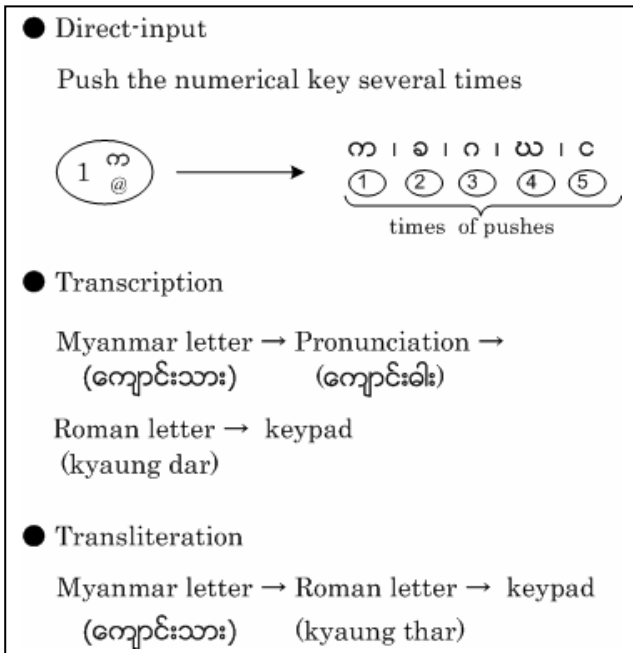


Fig.2 Approaches of Myanmar Language Input

5. Myanmar Romanization System

There is no standard Romanization system for Myanmar language so far. Many scholars have tried to unify those systems but none of them were successful system yet. For Romanization of Myanmar language, it is believed that there is no way to show accurate pronunciation of a word as well as its spelling in

Table.1 Comparison among Romanization Methods

Romanized Word	Romanization Method
sutesana	Duroiselle's System
thöötéthanã	Latter's System
thu.te-thãna.	Grant Brown's System (Conventional)
thúteθaná	Stewart's System (IPA)
thu.tei-thana.	Cornyn's System (typewritten)
thutéithãna	Minn Latt's 1966 System

Myanmar script. Table.1 is a comparison among existing Romanization methods for Myanmar word “သုတေသန” (research). [1]

As there is no standard Romanization system, the Romanized Myanmar spelling may differ depends on user. For example, the word “ကိုယ်” can be spelled as “ko” and it can be also spelled as “kol”. On the other hand, the spelling “ko” is also possible for the word “ကို” which has another possible spelling “go”.

In Myanmar language, many words are spelled differently from the way they are pronounced. For example, the word for “student” is pronounced as kyaung-dar (ကျောင်းခါး) but spelled as kyaung-thar (ကျောင်းသား).

Because of the above ambiguities, there should be a system that makes possible for the user to choose the right spelling of word. User may input the syllable in more than two ways of Romanized Myanmar and the system will have an intelligence to show a list of possible candidates in the selection menu.

6. Concept of the Romanized Myanmar Input

The concept is to input the pronunciation of Myanmar syllable, that user is thinking, in Roman script. Then the system will go and match the appropriate Myanmar syllable from the database. If the system finds out the corresponding syllable or syllables, it will show all the possible Myanmar syllables in the selection menu as a list of candidates. So that, user is able to select the right combination from the candidates appear on the selection menu as shown in the Fig.3.

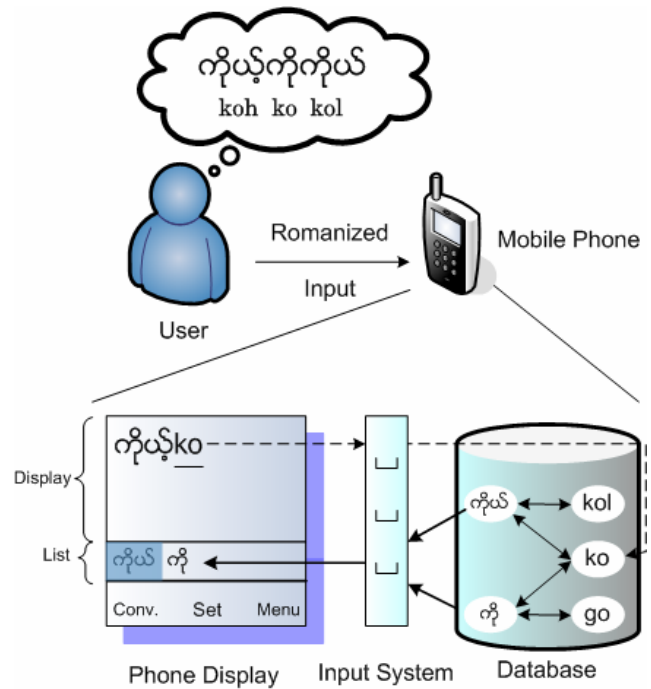


Fig.3 Conversion concept of Romanized Myanmar Input

6.1 Database for possible Romanized Myanmar Syllables

In order to develop Romanized Myanmar input method, database for the possible Romanized Myanmar syllables is Essential. Because natural language processing research for Myanmar language is not very developed, that kind of database is unavailable. Firstly, it is necessary to make a survey to get most frequently used Myanmar syllables. 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, etc. Total number of syllables used for this survey is 42,697. The result in frequency is shown in the Fig.4 with 1380 types of syllables.

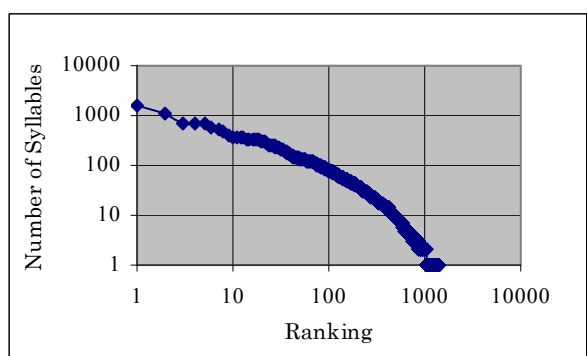


Fig.4 Frequency of Myanmar Syllables

6.2 Questionnaire Survey for most frequently used syllables

As the Romanized input of Myanmar Language varies by user, questionnaire survey for how to input those frequently used syllables is needed. For the Survey, 300 syllable types which is equivalent to 85% (approximately 30,000) of the total number of syllables are selected. Top twenty of the frequently used Myanmar syllables are shown in the Table 2 with possible Romanized Myanmar inputs. The Survey is continuing on 50 native Myanmar people from different areas and different education backgrounds. The questionnaire is created with the columns to fill sex, age, region, educational level, English proficiency and the syllables table.

7. Discussion and Conclusion

In this paper, we focus on Romanized input method for Myanmar language and when all the questionnaires are collected, the analysis on how people input Myanmar word in Roman script is going to be proceeding. By viewing those results, the average pattern of Myanmar writing is expected to be identified. Then we can finalize the database for the Romanized Myanmar syllables. In the near future, we are going to develop the Romanized Myanmar Input system for mobile phones

Table.2 The example of the questionnaire surveys

Frequently used Syllables	Romanized Myanmar	Frequency	Rank
အဲ	e, ee	1623	1
မ	ma	712	2
ကို	ko, go	710	3
က	ka, ga	704	4
သည်	thi, the	562	5
ရ	ya, ra	468	6
တ	ta	381	7
တယ်	te, tel	377	8
နေ	nay, nae	376	9
ဝါ	par, bar	369	10
တွေ	twae	357	11
ဖြစ်	phyit	339	12
ဆို	so, tso	338	13
မှ	mhar, mar	337	14
တော့	taw, tawt	329	15
သ	ta, tha	327	16
တဲ့	te, de	318	17
တ	tar, dar	301	18
လူ	lu, luu	293	19
များ	myar	292	20

that are used widely in Myanmar (for example, Nokia and Samsung). We have a plan to study about other famous input methods such as predictive conversion, and T9 input models in order to make more efficient input system.

References

[1] Ye Kyaw Thu and Yoshiyori Urano, Development of Romanized Input Methods for Myanmar Language Short Messaging Service (SMS): Comparison with Multitap Input Method, Simulation and Analysis, IPSJ SIG Technical Reports, vol.2006, no.3 (HI-117), pp.33-40.

[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.