Transcribing Chinese into Myanmar (Burmese) Script

†Chenchen Ding
Dept. of Computer Science
University of Tsukuba
1-1-1 Tennodai, Tsukuba,
Ibaraki, 305-8573, Japan
tei@mibel.cs.tsukuba.ac.jp

‡Ye Kyaw Thu, Eiichiro Sumita
†Multilingual Translation Lab
NICT
3-5 Hikaridai, Seikacho, Soraku-
gun, Kyoto, 619-0289, Japan
yekyawthu@nict.go.jp
eiichiro.sumita@nict.go.jp

Yoshinori Sagisaka
‡Dept. of Applied Mathe-
matics, GITI
Waseda University
3-4-1 Okubo, Shinjuku,
Tokyo, 169-8555, Japan
ysagisaka@gmail.com

Abstract

We propose a transcription system for Chinese into Burmese script. We design the system based on the phonology of Chinese and the orthography of Myanmar, referring other established Chinese transcription systems, i.e. the Pinyin system and the Palladius system. The Pinyin system transcribes Chinese into the Latin alphabet and the Palladius system transcribes Chinese into the Cyrillic alphabet. Our aim is to provide a system transcribing Chinese into the Burmese script. We hope our work may promote the cultural exchange of China and Myanmar.

1. Introduction

Linguistically, transcription is the systematic representation of language in written form. Different from transliteration, which is the conversion of a text from one script to another, transcription establishes a mapping between sound and script.

The transcription for Chinese[1] is especially an important issue because Chinese characters are logograms rather than phonograms. So, a transcription system from Chinese into a specific script will help the Chinese learners who use the script, as well as introduce Chinese concepts into the languages which adopt the script.

Several standard transcription systems for Chinese have been designed and applied. Hanyu Pinyin (Pinyin) [1] is the official phonetic system for transcribing Chinese into the Latin alphabet in China. Palladius system [2] transcribes Chinese into the Cyrillic alphabet, which is the Russian official standard for transcribing Chinese into Russian.

In this study, we design a transcription system for Chinese into Burmese script. The system is based on an investigation of the phonology of Chinese and the orthography of Myanmar. We try to design a system representing the pronunciation of Chinese as precise as possible, as well as keeping the consistency with conventional Myanmar orthography.

2. Transcription System

2.1. Chinese Syllable Structure

In this section, we first introduce the structure of Chinese syllables, and then describe the pro-

---

[1] In this paper, the expression of "Chinese" is always referred to the modern standard Chinese (i.e. Mandarin or Putonghua).

posed transcription system of different components within the syllable structure.

Syllables in Chinese have the maximal form of $CGVX^T$, where C is the initial consonant; G is one of the glides [i], [u], [y]; V is a vowel; X is a coda which may be one of [n], [ŋ], [i], [u]; and T is the tone. Any of C, G and X may be absent. C is called the "initial", G the "medial", and VX^T the "final". We show an example in Fig. 1 to illustrate the structure of a Chinese syllable.
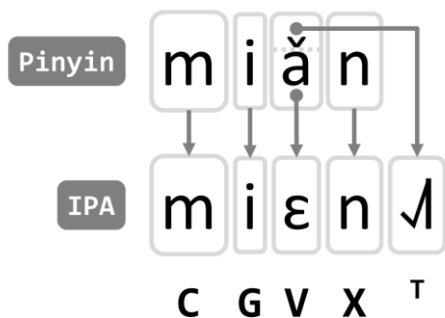

Figure 1. A Chinese syllable "miǎn".

We describe the transcription of "C" and "GVX" parts respectively in the next sections. For brevity, we refer the "GVX" part as "rhyme" or "final rhyme" in this paper.

2.2. Consonant Transcription

We show the transcription table of initial consonant in Table 1. The Chinese consonants are represented by Pinyin and the corresponding Burmese scripts are represented by the basic form of consonant alphabets. The International Phonetic Alphabet (IPA) transcription of Pinyin and Burmese alphabets are also listed. [2]

---

[2] The IPA transcription of the inherent vowel of Burmese consonant alphabets is omitted.

Table 1.   Initial consonant transcription.

Pinyin	Burmese	IPA	
		Chinese	**Myanmar**
-/y/w	အ	[ʔ]	[ʔ]
b	ပ	[p]	[p]
p	ဖ	[pʰ]	[pʰ]
m	မ	[m]	[m]
f	ဖ	[f]	[pʰʰ]
d	တ	[t]	[t]
t	ထ	[tʰ]	[tʰ]
n	န	[n]	[n]
l	လ	[l]	[l]
g	က	[k]	[k]
k	ခ	[kʰ]	[kʰ]
h	ဟ	[x]	[hʰ]
zh/j	ကျ	[tʂ]/[tɕ]	[tɕ]
ch/q	ချ	[tʂʰ]/[tɕʰ]	[tɕʰ]
sh/x	ရှ	[ʂ]/[ɕ]	[ɕ]
r	ရ	[ʐ]	[j]/[ɹ]
z	ဇ	[ts]	[z]
c	သ	[tsʰ]	[sʰ]
s	စ	[s]	[s]

Generally speaking, the stop and nasal consonants have satisfied correspondence between Chinese and Myanmar, while the fricative and affricate consonants have not. This is because Chinese has abundant fricative and affricate consonants while Myanmar has not. The details of Table 1 are described as follows.

1. Phonemes of [p], [pʰ], [m], [t], [tʰ], [n], [l], [k], [kʰ], [s] and empty consonant [ʔ] in Chinese have perfect correspondence in Myanmar phonemes.
2. Due to the absence of [f] in Myanmar, we add the "ha-hto" to "ဖ" to indicate the fricativeness.

3. Due to the absence of [x] in Myanmar, we add the "ha-hto" to "ဟ" to indicate the fricativeness. Actually, using [h] instead of [x] would not introduce confusion because [h] is not a phoneme in Chinese. We intend to emphasize the fricativeness of this phoneme.

4. Palatal phonemes [tɕ], [tɕʰ], [ɕ] in Chinese have their correspondence in Myanmar phonemes. However, the retroflex phonemes [tʂ], [tʂʰ], [ʂ] in Chinese can never find their correspondence in Myanmar. We do not distinguish the transcription of palatal and retroflex in our system. However, they can be phonotactically distinguished by their succeeding vowels, which we describe in the next section.

5. The palatal phoneme [ʐ] is transcribed into "ရ", considering "ရ" can be pronounced as [ɹ] in foreign words in Myanmar.

6. Phonemes of [ts] and [tsʰ] are transcribed to "ဇ" and "ဆ" separately. The transcriptions are not correct though; they are not confusing and can be distinguished from [s].

2.3. Rhyme Transcription

We show the transcription table of final rhymes of Chinese in Table 2.[3] They are represented by Pinyin in Chinese and by the glottal stop "အ" for corresponding Burmese transcription. The Burmese script of initial consonants listed in Table 1 can replace the first "အ" in the Burmese script of final rhymes listed in Table 2, to compose complete transcription of Chinese syllables. The IPA transcription of Pinyin and Burmese alphabets are also listed in Table 2.[4] The details of Table 2 are described as follows.

---

[3] The special vowel "er" ([ɑɻ]) is not listed in Table 2 (referring the detailed description #7). We do not consider the rhotic coda (erhua) in our system.

[4] The IPA transcription of glottal stop "အ" is omitted.

Table 2. Final rhyme transcription.

Pinyin	Burmese	IPA	
		Chinese	**Myanmar**
-i	အ(လ)	[ɻ̩]/[ɹ̩]	[u(l)]
a	အာ	[a]	[a]
e	အယ်	[ɯʌ]	[ɛ]
ai	အာအီ	[ai]	[ai]
ei	အေအီ	[ei]	[ei]
ao	အာအု	[ɑu]	[au]
ou	အိအု	[ou]	[ou]
an	အန်	[an]	[aɴ]
en	အိန်	[ən]	[eiɴ]
ang	အောင်	[ɑŋ]	[auɴ]
eng	အုင်	[ɤŋ]	[uɴ]
i	အီ	[i]	[i]
ia	အီအာ	[ia]	[ia]
ie	အီအယ်	[iɛ]	[iɛ]
iao	အီအာအု	[iɑu]	[iau]
iu	အီအီအု	[iou]	[iou]
ian	အီအိန်	[iɛn]	[iaiɴ]
in	အီအင်	[in]	[iiɴ]
iang	အီအောင်	[iɑŋ]	[iauɴ]
ing	အီအုန်	[iŋ]	[iuɴ]
u	အု	[u]	[u]
ua	အုအာ	[ua]	[ua]
uo/o	အုအော	[uɔ]/[o]	[uɔ]
uai	အုအာအီ	[uai]	[uai]
ui	အုအေအီ	[uei]	[uei]
uan	အုအန်	[uan]	[uan]
un	အုအိန်	[uən]	[ueiɴ]
uang	အုအောင်	[uɑŋ]	[uauɴ]
ong	(အု)အုန်	[uɤŋ]/[ʊŋ]	[(u)ouɴ]
ü	ယု	[y]	[ju]
üe	ယုအယ်	[yœ]	[juɛ]
üan	ယုအင်	[yɛn]	[juaiɴ]
ün	ယုအင်	[yn]	[juin]
iong	ယုအုန်	[iʊŋ]	[juun]

1. [a], [ai], [ei], [ɑu], [ou], [i], [ia], [iɛ], [iɑu], [iou], [u], [ua], [uɔ], [uai], [uei] can be relatively correctly transcribed due to the involved vowel phonemes are approximately similar in Chinese and Myanmar. We merge the "uo" and "o" in Pinyin for they are allophones and the merging will not lead to confusion.

2. [ɯʌ] in Chinese is very difficult to be transcribed into Myanmar. Actually, vowels of [ɯʌ] and [uɔ] compose a rounded/unrounded minimal pair. The "အယ်" we choose for [ɯʌ] is a quite different phoneme [ɛ] in Myanmar. The vowel is also used in the Palladius system for transcribing Chinese into the Cyrillic alphabet (i.e. "э"). We show the relation of the involved three phonemes in Fig. 2.

3. For both of the nasal codas: [n] and [ŋ] in Chinese, we use the nasal ending [N] in Myanmar to transcribe. A problem is that the [N] ending in Myanmar is a nasal which does not have a stable place of articulation. So, to clearly differentiate [n] and [ŋ] of Chinese is difficult. In the Palladius system, an interesting scheme is used that the [n] is transcribed using "нь" ([nʲ]) and the [ŋ] is transcribed using "н" ([n]). This is reasonable because [n] and [ŋ] of Chinese also affect their preceding vowels that before [n], the vowels tend to be front and before [ŋ], they tend to be back. So the Palladius system uses the soft sign "ь" to distinguish. We also refer to the scheme, using [-uN] phonemes in Myanmar to transcribe [ŋ], and [-iN] phonemes to transcribe [n] with preceding non-back vowels.

4. For "weng" only, we use "အွအင့်". For "ong", we use "အုင့်". E.g., "dong" will be "တုင့်".

5. For glide [y], we use "ယု" ([ju]) and for [yœ], we use "ယုအယ်" ([juɛ]), referring the same choices in Palladius system (i.e. "ю" and "юэ", respectively).

6. For the special vowel "-i" ([ɻ̩]/[ɹ̩]), we use "အိ(ဝ်)". The phoneme is generally difficult for foreigners. Especially, it always appears after the fricative and affricate consonants of [tʂ], [tʂʰ], [ʂ], [ʐ], [ts], [tsʰ] and [s], which are also the difficult consonants in our system. We select the Burmese script with the help of Myanmar native speakers and consider the common usage of Burmese script.

7. For the special vowel "er" ([ɑɻ]), we use "အာ(ဝ်)". Also, we select the Burmese script with the help of Myanmar native speakers.
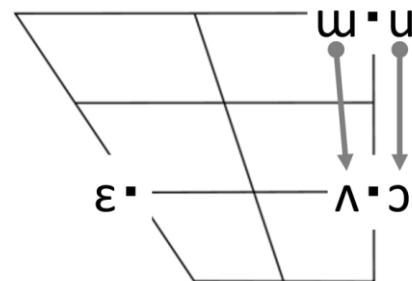

Figure 2. [ɯʌ], [uɔ], and [ɛ] in the IPA vowel chart. [ɯʌ] is unrounded and [uɔ] is rounded. [ɛ], [ʌ], and [ɔ] are all open-mid vowels.

2.4. About the Tones

We have not designed the transcription of tones in Chinese. Our system just employs the low tone in Myanmar. As Chinese and Myanmar are both tonal languages, to develop a tone mapping of the two languages seems possible. However, the tones are not matching well in the two languages. Further, there is complicated sandhi of tones in Chinese, where the tones will change depending on the context. In this study, we do not get into the topic of tone transcription and leave it in our future studies.

3. Experiment

3.1. Automatic Transcription

We have developed a program for automatically applying the proposed transcription system. The process is along a line of "Chinese character → Pinyin → Burmese script", as the following example shows.

其实地上本没有路，走的人多了，也便成了路。

↓

qi shi di shang ben mei you lu, zou
de ren duo le, ye bian cheng le lu.

↓

ချိ ရှ(လ်) တိ ရှောင် ပိန် မေအီ အီအို�-အူ လူ ဇိုအူ တယ်
ရိန် တူအော် လယ်၊ အီအယ် ပိအိုင် ချုန် လယ် လူ။

For the Chinese character-to-Pinyin step, we used a Python-based toolkit `pypinyin` [3] with the `jieba` [4] Chinese segmentation tool. Both of the tools are open-sourced. A problem in this step is that some Chinese characters have multiple pronunciations dependent on the context. However, the tools we used could give a relatively satisfied performance after we trimmed several bugs in them.

For the Pinyin-to-Burmese script step, we created the mapping table for all possible Chinese syllables. Generally, the mapping system from Chinese character to Burmese script is not complex as long as we do not consider the tones.

3.2. User Study Settings

In order to examine whether the proposed transcription system can represent Chinese pronunciations correctly, we conducted a user study. The user study was designed as follows.

1. We selected Chinese texts and used our program to transcribe them in Burmese script automatically[5]. The Chinese texts were selected from the novels and essays of the famous Chinese writer Lu Xun [5][6] and segmented to short clauses. Finally we had a text set of 116 lines with an average length around 10 syllables. The text set contains all the consonants and rhymes in Chinese phonology as listed in Table 1 and Table 2.[7] The syllables contained in the text set also cover approximate half of all the possible Chinese syllables.[8]

2. Five Myanmar native speakers[9] were asked to read the prepared text by the transcribed Burmese script. Their reading was recorded. Before their reading, all of them had been familiarized with the transcription system by hearing a Chinese native speaker's[10] reading of all the possible Chinese syllables with an introduction of the transcription system.

3. Four Chinese native speakers[11] were asked to hear the Myanmar native speakers' reading. They were asked to give a point from 1 to 7, to tell whether the pronunciation was heard as natural, standard Chinese. 7 means perfect and 1 means poor. They were also asked to point out syllables with unsatisfied pronunciation. The examination is conducted line by line. For the Chinese native speakers, a piece of recording was played for each line randomly selected from the recordings of the five Myanmar native speakers' reading.

---

[5] We also did manual check to avoid mapping errors.

[6] Excerpts of *一件小事*, *故乡*, and *藤野先生*.

[7] Including the vowel "er" but not rhotic coda.

[8] The total types of Chinese syllables are different from literatures due to marginal ones. Generally, there are around 400 types of syllables disregarding the tones. Our text set contains around 200 types.

[9] Four females and one male (the second author).

[10] Male (the first author).

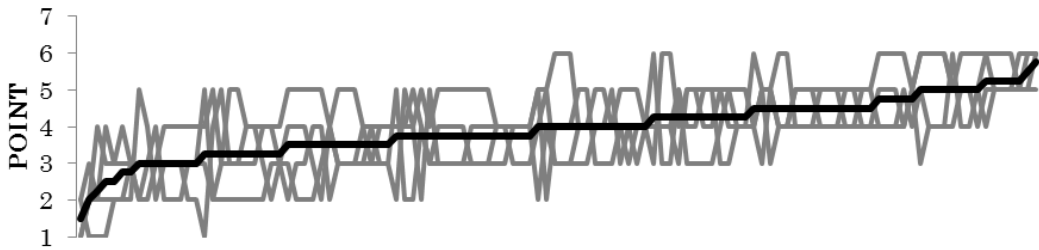[11] All males, knowing nothing about the transcription.

Figure 3. Average points for all lines (black) in our text set and the points from different evaluators (grey). Here the lines are sorted by their average points. Lines with a low average (left side) have relatively larger variances among evaluators than those with a high average (right side).

3.3. Evaluation Results

First, we show the points for all the lines in our text set, given by the evaluators, i.e. the four Chinese native speakers (Fig. 3). We can observe that the accordance among the evaluators is very rough, that the grey lines do not obviously converge to the black line in Fig. 3. The phenomenon suggests that the four evaluators may have different preference in the evaluation. We list the average and variance of the evaluators' points in Table 3. The averages are around 4 (out of 7) points, which illustrate a mediocre performance of our system. Obviously, evaluator #2 has a larger variance and #4 has a higher average.

Table 3.　Average and variance of evaluators.

evaluator	#1	#2	#3	#4
average	3.7	3.8	3.6	4.7
variance	1.1	1.6	0.7	0.9

In Fig. 4, we give a further investigation of the evaluation results. We observe that 3/4 of the total lines (87/116) have a variance less than 1.0 among different evaluators. We further calculate the average and variance of the 87 lines and list the results in Table 4. We observe the averages increase (except #4, whose original average is high) and variances decrease (except #3, whose original variance is low) on this sub-set. We consider that an agreement may be more difficult to achieve among different evaluators, on those lines with relatively unsatisfied pronunciations.
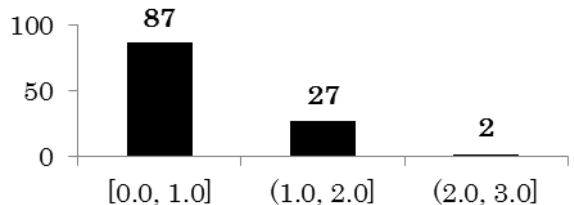

Figure 4.　Distribution of variance among evaluators for all the lines.

Table 4.　Average and variance of lines with a variance less than 1.0 among evaluators.

evaluator	#1	#2	#3	#4
average	3.8	3.9	3.7	4.6
variance	1.0	1.4	0.7	0.8

4. Discussion

In this section, we conduct a further investigation of the performance on specific phonemes in our transcription system. As we have asked the evaluators to mark the unsatisfied pronunciation of specific syllables, we have statistics of the problematic consonants and rhymes. The analysis is based on the common problematic phonemes among different evaluators.

Pinyin	Burmese	Types of unsatisfied syl.			
		#1	#2	#3	#4
b	ဗ	2	0	3	3
p	ပ	2	2	3	1
m	မ	1	1	2	2
f	ဖ	2	0	1	0
d	တ	4	1	5	1
t	ထ	4	0	2	4
n	န	1	0	1	0
l	လ	3	3	4	5
g	ဂ	2	1	2	2
k	ခ	1	0	0	0
h	ဟ	6	7	5	4
zh/j	ကျ	13	8	8	9
ch/q	ချ	5	3	3	6
sh/x	ရှ	12	5	8	8
r	ရ	1	4	3	3
z	ဇ	5	3	3	4
c	ဆ	5	3	6	6
s	စ	1	0	1	2

Table 5.   Unsatisfied consonants

Pinyin	Burmese	Types of unsatisfied syl.			
		#1	#2	#3	#4
-i	အ(လ်)	5	3	5	5
a	အာ	0	0	0	0
e	အယ်	4	4	6	3
ai	အာအိ	3	2	1	2
ei	အေအိ	1	0	0	0
ao	အာအု	3	2	2	0
ou	အိအု	4	1	1	1
an	အန်	3	2	7	4
en	အိန်	1	1	3	2
ang	အောင်	4	4	4	5
eng	အွန်	6	2	5	5
i	အိ	3	1	2	2
ia	အိအာ	1	0	0	0
ie	အိအယ်	2	1	2	1
iao	အိအာအု	2	0	0	0
iu	အိအိအု	3	1	1	2
ian	အိအိင်	3	2	3	7
in	အိအင်	3	1	1	0
iang	အိအောင်	2	3	3	3
ing	အိအွန်	3	1	2	3
u	အု	7	0	3	2
ua	အုအာ	0	0	0	0
uo/o	အုအော်	2	0	1	3
uai	အုအာအိ	0	1	1	0
ui	အုအေအိ	3	1	1	1
uan	အုအန်	0	2	0	2
un	အုအိန်	3	3	1	4
uang	အုအောင်	2	2	0	0
ong	(အု)အန်	3	1	5	3
ü	ယု	2	4	2	4
üe	ယုအယ်	4	1	2	1
üan	ယုအိင်	1	2	1	1
ün	ယုအင်	1	0	0	0
iong	ယုအွန်	1	0	0	0

Table 6.   Unsatisfied rhymes.

We count the types of unsatisfied syllables containing specific consonant and rhyme and list them in Table 5 and Table 6. The absolute value of the types is meaningless, because different consonants and rhymes have different frequencies in texts and different evaluators have different tolerances. From Table 5 and Table 6, it is obviously that evaluator #1 tended to mark more and #2 tended to mark less. We thus check the different counts of types and take the high half as problematic ones. For example, in Table 6, #2 has five different counts of types as 0, 1, 2, 3 and 4, so we take the rhymes with the two (no more than half of five) counts of 3 and 4 as problematic ones. We mark them by grey cells. For those common problematic consonants and rhymes, we also mark the corresponding Pinyin and Burmese script by grey of different thicknesses.

From Table 5, we observe that the evaluators achieved agreement in most of the consonants. As we have described, all the stop and nasal consonants can be mapped relatively well but not for the fricative and affricate ones. For fricative consonants, only the "s - စ" pair has a satisfied performance because it is the only pair can be accurately mapped between Chinese and Myanmar. As to the most unsatisfied consonant mapping, the pairs of "h - ဟျ", "zh/j - ကျ", "sh/x - ရှ" and "c - ဆ" are marked by over half of the evaluators. From theories as well as our experiment, the fricative and affricate consonants are the most difficult parts in the transcription system.

From Table 6, we observe the common problematic ones are pairs of "-i - အꩥ(လ်)", "e - အယ်", "ang - ေအာင်" and "eng - အွꩥ်". The former two pairs, as we have described, are really difficult for mapping[12]. For the latter two pairs, we find the [-uɴ] ending in Myanmar is usually pronounced very light, so the final nasal may be inaudible for Chinese listeners because the [ŋ] ending in Chinese is a very "heavy" nasal. So for the evaluators, "ang" was sometimes heard as "ao". The problematic pair "iang - အီေအာင်" in #2 can be attributed to the same reason.

We also notice different preferences of evaluators. For example, #1 is more unsatisfied with the "u - အꩥ" pair than the other three. Because the roundness of the vowel [u] is greater in Chinese than in Myanmar, #1 may be more sensitive on the roundness. We consider this is the same reason for the "ou - အꩥ̊အꩥ" pair becomes a problematic one for #1. Another example is the "ian – အီအꩥင်" pair for #4. Noticing #4 is also relatively unsatisfied with "an - အန်" and "un - အꩥအန်"

pairs, we consider #4 may be more sensitive on the backness of vowels before the nasal [n].

5. Conclusion and Future Work

In this study, we proposed a transcription system for Chinese to Burmese script. We designed it according to Chinese phonology, and examined it by user study. The system reaches moderate performance considering the phonological difference between the two languages.

In future work, we first plan to conduct deeper investigation of the unsatisfied pronunciations revealed in our experiment, in theories and in practice. We also plan to develop our transcription system to include the mapping of tones between Chinese and Myanmar.

Acknowledgment

References

[1] "汉语拼音方案" (Scheme of the Chinese Phonetic Alphabet), adopted on 1958.02.11 by the National Assembly of the People's Republic of China.
[2] S. A. Matsaev and V. G. Orlov, "Пособие по транскрипции и правописанию китайских слов" (Manual transcription and spelling Chinese words), Moscow, 1966.
[3] http://pypinyin.readthedocs.org/en/latest/
[4] https://github.com/fxsjy/jieba
[5] http://zh.wikisource.org/wiki/Author:%E9%AD%AF%E8%BF%85

---

[12] The "er - အာ(လ်)" pair is satisfied by evaluator #1 and #2, but not by #3 and #4.