

Transcription of Pure Korean Proper Nouns into Chinese Characters

Jeongmyeong Kim¹⁺ and Taeseon Yoon²

Hankuk Academy of Foreign Studies

Abstract. For the strong ideographic characteristic of Chinese characters, translation into Chinese Language often requires the process of transcription in which irreplaceable words from foreign origins such as foreign proper nouns are expressed in Chinese characters that convey only the sounds of the characters, not the meaning. In the case of foreign languages in Sino-spherical boundaries, such as Korean and Japanese, words with the origin of Chinese characters are translated in the equivalent simplified Chinese characters. However, with the recent decrease of the usage of Chinese characters in the creation of proper nouns in Korean language, confusion arouses with the absence of any set criteria for properly translating or transcribing Korean proper nouns using pure Korean words into Chinese characters. This paper, in recognition of the strong need to establish the transcription rule for pure Korean proper nouns, intends to find the rule applied in the standard of Chinese transcriptions and adopt the same rule to newly establish the transcription rule for Korean language. For the pronunciations particular in Korean language, the paper makes use of Hidden Markov Model (HMM) to determine the most proper transcription, considering International Phonetic Alphabet.

Keywords: Transcription, Pure Korean Proper Nouns, Hidden Markov Model (HMM).

1. Introduction

Since Chinese characters are ideographic, translations of irreplaceable foreign vocabularies usually resort to the use of transcription. In Republic of Korea, since Korean language is in the boundary of Sino-sphere, usage of both Korean alphabets and Chinese characters had been historically pervasive throughout the nation until the government officially recognized Korean alphabet as the primal letters in the writing of all official documents in 1970. The use of Chinese characters has since rapidly reduced not only in documents and textbooks but also in daily lives of Korean people. Korean proper nouns had been easily translated into Chinese because most of them, including names of people and corporations, had had their origin in Chinese characters in general. However, with the decrease in the usage of Chinese characters, the use of pure Korean proper nouns has vastly increased. According to the official criteria set by Seoul city on the translation of Korean language into Chinese[6], names of places in pure Korean are to be translated in consideration of their original meanings. However, when the meaning or origin of the name is too vague to be translated or its historical appellation in Chinese language is considered too different from its Korean name now, the name is transcribed into Chinese characters. People's Republic of China has set the standard of transcription of foreign language in *Names of the World's Peoples*[3], but the book does not include the transcription of languages in Sino-sphere, including Korean and Japanese. Therefore, when pure Korean proper nouns are transcribed into Chinese characters, they are either transcribed with Romanization (김미나->金 Mina) or arbitrarily transcribed into Chinese characters by the translator (김미나->金米娜). Since the standard for the translation of these proper nouns is set by neither Chinese government nor Korean government yet, the problem may cause further confusion in Chinese transcription of Korean language. In this paper, I confine the object of transcription solely to pure Korean proper nouns. I exclude pure Korean vocabularies that have equivalence with Chinese words. For instance, '뫼' is a pure Korean noun meaning 'mountain'. As the vocabulary is equivalent to Chinese word '山', the word is not subject to transcription. Moreover, I exclude proper nouns that are not purely originated from Korean, possessing corresponding Chinese characters from its origin. A typical Korean name '김정명' is a Korean proper noun. However, since it has the origin in Chinese characters '金正明', it is not subject to transcription.

⁺ Corresponding author. Tel.: + 821090691160; fax: +82313225825.
E-mail address: aud521@naver.com.

2. Relative Research

The related researches are mostly about the translation between Chinese and Korean or among East Asian Sino-spherical languages. NTCIR Workshop in Tokyo[5] researched on cross-lingual information retrieval among Korean, Chinese, and Japanese. Another research from the workshop[7] focused on the system of translation between Korean and Chinese. My research has to do with the translation system because the transcription rule can be applied for mechanical translation.

Naito Eisuke[1] concentrates on the authority control. The name of a person is described into different script systems, and variant forms or variations are drastically increasing, and creating a lack of integrity for retrieval precision. This relates to the purpose of my paper that the standard transcription rule should be established even for Sino-spherical languages in its translation to Chinese language.

Fujitsu Research and Development Centre[2] reported a research on effective translation of Chinese vocabulary to English based on Semantic Prediction. However, since proper nouns should not be translated but transcribed in terms of phonology, the research cannot address the problem in transcription, although it can be applied for translation of Chinese terms into Korean or vice versa through semantic prediction.

The research on the transcription of Korean language into Chinese is quite unprecedented, with neither Chinese nor Korean government offering a standard rule and few researches focusing on the transcription of Korean into Chinese.

3. Experiment

3.1. Design

<Table 1> Transcription of Korean into Chinese

	ㅂ [p] ㅃ [pʰ] ㅍ [p]	ㅍ [pʰ]	ㅁ [m]	ㄷ [t] ㅌ [tʰ]	ㅌ [tʰ]	ㅅ [s]	ㅆ [ʃ]	ㄴ [n]	ㄹ [r/l]	ㅈ [tɕ]	ㅊ [tɕʰ]	ㅋ [k] ㆁ [kʰ]	ㆁ [kʰ]	ㅎ [h]	*ㅇ	
ㅏ [a]	巴/芭 bā/bā	帕 pà	马/玛 mǎ/mǎ	达 dá	塔 tǎ	沙/莎 shā/shā	萨 sà	纳/娜 nà/nà	拉 lā	贾 jiǎ	扎 zhā	察 chá	加 jiā	卡 kǎ	哈 hā	阿 ā
ㅓ [jɛ]	比亚 bǐyà	皮亚 píyà	米亚 mǐyà	迪亚 dìyà	蒂亚 dìyà	希亚 xīyà	西亚 xīyà	尼亚 nǐyà	利亚 lìyà	吉亚 jíyà	贾 jiǎ	恰 qià	吉亚 jíyà	基亚 jīyà	希亚 xīyà	亚 yà
ㅗ [ʌ]	伯 bó	珀 pò	默 mò	德 dé	特 tè	舍 shě	瑟 sè	纳/娜 nà/nà	勒 lè	哲 zhé	泽 zé	策 cè	格 gé	克 kè	赫 hè	厄 è
ㅛ [jʌ]	比厄 bǐè	皮厄 píè	米厄 mǐè	迪厄 dìè	蒂厄 dìè	希厄 xīè	西厄 xīè	尼厄 nǐè	利厄 lìè	吉厄 jíè	齐厄 qíè	齐厄 qíè	吉厄 jíè	基厄 jīè	希厄 xīè	伊厄 yīè
ㅜ [o]	博 bó	波 bō	莫 mò	多 duō	托 tuō	肖 xiāo	索 suǒ	诺 nuò	洛 luò	乔 qiáo	佐 zuǒ	措 cuò	戈 gē	科 kē	霍 huò	奥/欧 ào/ōu
ㅠ [jo]	比奥 bǐào	皮奥 píào	米奥 mǐào	迪奥 dìào	蒂奥 dìào	希奥 xīào	肖 xiāo	尼奥 nǐào	利奥 lìào	吉奥 jíào	齐奥 qíào	齐奥 qíào	吉奥 jíào	基奥 jīào	希奥 xīào	约 yuē

ㅏ [u]	布 bù	普 pǔ	穆 mù	杜 dù	图 tú	舒 shū	苏 sū	努 nǔ	卢 lú	朱 zhū	祖 zǔ	楚 chǔ	古 gǔ	库 kù	胡 hú	乌 wū
ㅑ [ju]	比 bǐ 乌 wū	皮 pí 乌 wū	缪 miù	迪 dí 乌 wū	蒂 dì 乌 wū	希 xī 乌 wū	休 xiū	纽 niǔ	柳 liǔ	吉 jí 乌 wū	齐 qí 乌 wū	齐 qí 乌 wū	吉 jí 乌 wū	基 jī 乌 wū	希 xī 乌 wū	尤 yóu
ㅡ [w]	布 bù	普 pǔ	姆 mǔ	德 dé	特 tè	什 shí	斯/ 丝 sī /sī	恩 ēn	尔 ěr	奇 qí	兹 zī	茨 cí	格 gé	克 kè	赫 hè	乌 wū
ㅣ [i]/ ㅓ [ui]	比 bǐ	皮 pí	米 mǐ	迪 dí	蒂 dì	希 xī	西 xī	尼/ 妮 ní /nī	利/ 莉 lì /lì	吉 jí	齐 qí	齐 qí	吉 jí	基 jī	希 xī	伊 yī
ㅕ [ɛ]/ ㅖ [e]	贝 bè	佩 pè	梅 mé	德 dé	特/ 泰 tè /tài	谢 xiè	塞 sāi	内 nè	莱 lái	杰 jié	泽 zé	策 cè	盖 gài	凯 kǎi	赫/ 黑 hè /hēi	埃 āi
ㅗ [jɛ]/ ㅛ [je]	别 bié	皮 pí 耶 yē	米 mǐ 耶 yē	迭 dié	铁 tiě	谢 xiè	谢 xiè	涅 niè	列 liè	杰 jié	齐 qí 耶 yē	切 qiē	吉 jí 耶 yē	基 jī 耶 yē	希 xī 耶 yē	耶 yē
ㅜ [wɛ]	布 bù 阿 ā	普 pǔ 阿 ā	穆 mù 阿 ā	杜 dù 阿 ā	图 tú 阿 ā	舒 shū 阿 ā	苏 sū 阿 ā	努 nǔ 阿 ā	卢 lú 阿 ā	朱 zhū 阿 ā	祖 zǔ 阿 ā	楚 chǔ 阿 ā	瓜 guā	夸 kuā	华 huá	瓦 wǎ
ㅟ [wɛ]/ ㅠ [ø]/ ㅡ [we]	布 bù 埃 āi	普 pǔ 埃 āi	穆 mù 埃 āi	杜 dù 埃 āi	图 tú 埃 āi	舒 shū 埃 āi	苏 sū 埃 āi	努 nǔ 埃 āi	卢 lú 埃 āi	朱 zhū 埃 āi	祖 zǔ 埃 āi	楚 chǔ 埃 āi	圭 guī	奎 kuí	惠 huì	韦 wéi
ㅓ [wʌ]	博 bó	波 bō	莫 mò	多 duō	托 tuō	索 suǒ	索 suǒ	诺 nuò	洛 luò	佐 zuǒ	佐 zuǒ	措 cuò	果 guǒ	阔 kuò	霍 huò	沃 wò
ㅕ [qi]	布 bù 伊 yī	普 pǔ 伊 yī	穆 mù 伊 yī	杜 dù 伊 yī	图 tú 伊 yī	绥 suí	绥 suí	努 nǔ 伊 yī	卢 lú 伊 yī	朱 zhū 伊 yī	祖 zǔ 伊 yī	楚 chǔ 伊 yī	圭 guī	奎 kuí	惠 huì	维 wéi

*‘ ㅑ ’ here is the null consonant.

**Characters shown as alternatives above, written after ‘/’, can be used for female names.

Korean phonology consists of three allophones: initial, medial, and final. Since Mandarin Chinese phonology includes few final allophones, the paper will primarily concentrate on initial and medial allophones of both languages.

In order to base the rules of the new graph on the official transcription rule set for other languages in *Names of the World's Peoples*[3], transcription table for English, Spanish, and Portuguese were used. Furthermore, to correspond each Korean consonant or vowel to English, Spanish, and Portuguese, I referred to International Phonetic Alphabet (IPA) and the transcription rule set by The National Institute of the

Korean Language for the transcription of Chinese into Korean[4]. In terms of HMM[8], the priority is as follows: first, IPA and the standard of transcription of English, Spanish, and Portuguese in *Names of the World's Peoples*[3]; second, transcription rule set by The National Institute of the Korean Language[4]; last, the rule found in other languages. The steps are as follows:

First, I found the corresponding English, Spanish, and Portuguese consonants or vowels with Korean. The correspondence can be determined by the comparison of IPA. For instance, bilabial sound [ㅍ] corresponds with English consonant [p] because both consonants have the same IPA of [p^h]. Korean diphthongs were found corresponding to Spanish and Portuguese diphthongs mostly. [ㅟ] corresponds with Spanish vowel [ia] or [ya] with the same or similar IPA of [jɐ]. Most of the Korean phonology could find its corresponding Chinese transcriptions with the comparison with other language in this manner.

Second, if the corresponding consonants or vowels are difficult to be found, I referred to the transcription rule set by The National Institute of the Korean Language[4]. The rule is officially set for the proper transcription of Chinese language into Korean. For example, it is hard to figure out how to convey the difference between alveolar sound [ㄷ] and [ㄷ], or [s] and [s-], since none of English, Spanish, or Portuguese consonants contain the same or similar alveolar differences. In this case, according to the Korean transcription rule aforementioned, Chinese pinyin [s] is transcribed into [ㄷ] and [sh] into [ㄷ]. Therefore, transcription of [ㄷ] referred to the transcription of English [s] because [s] sound English consonants were mostly transcribed into Chinese characters with pinyin of [s] as their consonants. In the same way, transcription of [ㄷ] referred to that of English [ʃ].

Third, exceptional transcriptions followed the rule found in other languages. [ㅟ] could not find its corresponding English, Spanish, or Portuguese. However, just as [ia] sound in Spanish were divided into [i] and [ya] in its transcription, [ㅟ] could also be split into [i] part and [ə] part. the latter part used [ㅟ, ə].

3.2. Manuscript Requirements

The steps for transcribing Korean phonology into Chinese characters would be as follows, based on the graph above.

First, pure Korean proper noun is transcribed into Korean phonetic expression. Say that ‘푸름이’ is a pure Korean proper noun. ‘푸름이’ is transcribed as [푸르미].

Second, according to the transcription table, [푸] is transcribed as ‘普’, [르] as ‘尔’, and [미] as ‘米’.

Therefore, transcription of ‘푸름이’ is ‘普尔米’, or ‘pǔ ěr mǐ’.

4. Conclusion

The limitation of this research is that it could not focus on final allophones of Korean phonology. Since final allophones are integral parts of the Korean phonology, correspondence of these allophones to Chinese phonology should be researched further. Another limitation is that, although I exerted best efforts to convey both original Korean pronunciation and Chinese characters used in transcription tables for other languages, there may be some characters that could better be substituted with other characters to convey the original Korean phoneme more effectively.

Although the research on Chinese transcription has been abundant, the matter of transcription from Korean to Chinese has been rarely researched by now. With the increase in demand for proper transcription rule from Korean to Chinese, I hope that this research will serve as the underpinning of the establishment of the official transcription rule, for both Republic of Korea and the People’s Republic of China.

5. Reference

- [1] Eisuke, Naito. *Names of the Far East: Japanese, Chinese, and Korean Authority Control*. Thesis. Toyo University, n.d. N.p.: n.p., n.d. Web.
- [2] Fang, Gaolin, Hao Yu, and Fumihito Nishino. *Chinese-English Term Translation Mining Based on*. Thesis. Fujitsu Research and Development Center, Co., LTD. Beijing 100016, China, n.d. N.p.: n.p., n.d. Web.
- [3] Guo, Guorong. *Shi Jie Ren Ming Fan Yi Da Ci Dian / Names of the World's Peoples : A Comprehensive Dictionary of Names in Roman-Chinese / Compiled by Proper Names and Translation Service Xinhua New*

Agency. Bei Jing: Zhong Guo Dui Wai Fan Yi Chu Ban Gong Si, 1993. Print.

- [4] Kim, Jonghwan, ed. *Hankuk Eomun Gyujeongjib (Official Prescription for the use of Korean Grammars)*. Ed. Hoseong Jeong. 3rd ed. Seoul, ROK: The National Institute of the Korean Language, 2011. 131-32. Print.
- [5] Kishida, Kazuaki, Kuang-hua Chen, Sukhoon Lee, Hsin-Hsi Chen, Noriko Kando, Kazuko Kuriyama, Sung Hyon Myaeng, and Koji Eguchi. "Cross-Lingual Information Retrieval (CLIR) Task." Thesis. Tokyo, Japan, 2004. NTCIR Workshop (n.d.): n. pag. Web.
- [6] "Dictionary for the Official Transcription of Korean Language in Seoul City." *Dictionary for the Official Transcription of Korean Language in Seoul City*. The Metropolis of Seoul, 2013. Web. 08 Dec. 2013.
- [7] Wang, Yu-Chun, Cheng-Wei Lee, Richard Tzong-Han Tsai, and Wen-Lian Hsu. "IASL System for NTCIR-6 Korean-Chinese." Thesis. Tokyo, Japan, 2007. NTCIR-6 Workshop Meeting (2007): n. pag. Print.
- [8] Devijver, Pierre A. "Baum's forward-backward algorithm revisited." *Pattern Recognition Letters* 3.6 (1985): 369-373.