Linguistic Research 40(Special Edition): 33-59
DOI: 10.17250/khisli.40..202309.002

Orthographic influence on resyllabification errors by Vietnamese learners of Korean: A speech corpus study*

Hye-Won Choi · Gunhee Ko
(Ewha Womans University)

Choi, Hye-Won and Gunhee Ko. 2023. Orthographic influence on resyllabification errors by Vietnamese learners of Korean: A speech corpus study. *Linguistic Research* 40(Special Edition): 33-59. This study investigates resyllabification errors made by Vietnamese learners of Korean, utilizing an L2 speech corpus known as "A Collection of Foreigners' Korean Speech Data for AI Training." By analyzing the read-aloud and spontaneous speech data from 40 Vietnamese L1 speakers in the corpus, we discovered that the error rate was higher in script reading than in spontaneous speech. We propose the rigid syllable boundary in Vietnamese L1 phonology, contrasting with the fluid syllable boundary in Korean L2 phonology, results in a negative transfer, causing Vietnamese learners' difficulty in resyllabification when speaking Korean. Furthermore, the orthographic rigidity of syllable boundary of the Korean writing system presents negative orthographic input for Vietnamese learners, whose orthography (Roman alphabet) allows for syllabic ambiguity. Consequently, speakers tend to produce more errors in script reading with such orthographic input than in spontaneous speech. We conclude the disparity between the phonological fluidity and the orthographic rigidity of Korean syllables contributes to failure of resyllabification among Vietnamese speakers. **(Ewha Womans University)**

Keywords Korean as a foreign language (KFL), Vietnamese learners, resyllabification, second language pronunciation errors, orthography, speech corpus, syllable boundary

1. Introduction

As one might expect, second language learners often make pronunciation errors, and Vietnamese learners of Korean are no exception. Vietnam ranks third in terms of TOPIK

* We are grateful to two anonymous reviewers for their invaluable and insightful comments, which helped clarify and solidify some of the arguments of this paper. Any remaining errors or shortcomings are of course ours.

(Test of Proficiency in Korean) test applicants (TOPIK 2021) and the Vietnamese constitute the second largest foreign community in Korea (Ministry of Justice 2021).[1] As the size of the Vietnamese L1 speakers increases, more attention has been placed on various aspects of their L2 acquisition of Korean. Although the recent trend in second language acquisition has favored achieving fluency over accuracy, the accuracy of the target language pronunciation still contributes greatly to successful communication. Vietnamese learners have been reported to display various types of pronunciation errors: in segmental realization, confusing the three-way distinction (i.e., lenis, fortis, and aspirated) of Korean stops (Jang 2018a; Yoo and Kang 2018; Lee 2019) and affricates (Jang 2018b), omitting coda ㄹ [l] or replacing it with ㄴ [n] (Cho 2005; Hwang 2016), pronouncing diphthongs as a sequence of monophthongs (Cho 2005); and failing to apply phonological rules such as glottalization (Kim 2004), nasalization, palatalization, (Lee and Cho 2022), and resyllabification (also called "liaison") (Lee and Cho 2022; Lee and Kwon 2022). Generally, the application of phonological rules is known to require a deeper understanding of the language than segmental realization, since these rules are not evidently visible on the surface and thus unpredictable from the spelling.

It has long been discussed that the second language orthographic input has a considerable impact on the learners' phonological L2 production, both positively and negatively (Erdener and Burnham 2005; Bassetti 2006, 2008; Rafat 2011; Escudero et al. 2013; Han and Kim 2017; Cerni et al. 2019). On one hand, the L2 orthography may hinder learners from pronouncing target-like speech sounds, especially when L1 and L2 have different transparency levels in the writing system (Escudero et al. 2013). On the other hand, the L2 writing system can also help learners who have difficulty perceiving complicated L2 sounds or words, by complementing the acoustic input and their L2 perception defects and thus fostering target-like L2 speech production (Erdener and Burnham 2005; Bassetti 2008; Cerni et al. 2019).

Korean writing system Hangeul is known to be shallow or transparent. Transparent orthography refers to one that has a one-to-one correspondence between graphemes (spelling) and phonemes (sound). Opaque orthography, by contrast, indicates an inconsistent, one-to-multiple grapheme-to-phoneme correspondence. Thanks to its

1 The top 5 nations (excluding Korea), measured by the number of applicants to take the TOPIK (Test of Proficiency in Korean) test from 2017 to 2021, are China, Japan, Vietnam, Taiwan, and Thailand (TOPIK 2021), and the top 5 groups of foreigners living in Korea are Chinese, Vietnamese, Thai, American, and Uzbek (Ministry of Justice 2021).

relatively transparent orthography, Korean Hangeul is assessed to be one of the easiest scripts to learn and therefore there have been relatively few studies on learners' pronunciation errors caused by the orthography of Korean (cf. Lee 2012; Kim 2022).

This study investigates resyllabification errors as illustrated in (1), which is one of the frequent types of errors produced by Vietnamese learners of Korean.[2] While native speakers of Korean pronounce the coda of the preceding syllable as the onset of the following syllable by applying resyllabification as in the right column, Vietnamese L2 speakers of Korean often fail to apply resyllabification and thus tend to sound "choppy," as in the left column.[3] (The syllable boundary is marked by a . (dot) in the examples.)

(1) Resyllabification errors produced by Vietnamese L1 speakers[4]

	Vietnamese learners	Korean speakers
a. *halin* (할인) 'discount'	[ha**l**.in]	[ha.**ɾ**in]
b. *hankwuke* (한국어) 'Korean'	[han.ku**k**.ʌ]	[haŋ.gu.**g**ʌ]
c. *pom-ey* (봄에) 'in spring'	[po**m**.ɛ]	[po.**m**ɛ]
d. *pat-un* (받은) 'receive-Mod'	[pa**t**.ɯn]	[pa.**d**ɯn]
e. *Seoul eti-ey* (서울 어디에) 'where in Seoul'	[sʌ.u**l**.ʌ.ti.ɛ]	[sʌ.u.**ɾ**ʌ.di.ɛ]

Utilizing an L2 Korean speech corpus known as "A Collection of Foreigners' Korean Speech Data for AI Training" (Yoon 2021), we specifically examine if the L2 orthography influences the realization of resyllabification by comparing the error rates

---

2 We used Yale Romanization for the trasliteration of Korean examples in this paper, except for proper names such as Hangeul.

3 The resyllabification errors are exemplified in two different patterns by the Vietnamese L1 speakers. First, they pronounce the final consonant in its original coda position, not moving it into the onset of the next syllable as in (1). Second, some learners duplicate the coda sound and pronounce it twice, as below in (i). This second pattern is also classified as a type of resyllabification error in that the coda consonant stays in its original position, as is likewise classified in Lee and Kwon (2022).

(i) *kakyek-i* (가격이) 'price-Nom' [ka.gjʌ.gi] → [ka.gjʌk.gi]

4 The liquid sound ㄹ in Korean has two phonetic realizations: [l] as coda and [ɾ] as onset. Therefore, when the coda ㄹ becomes the onset of the next syllable by being resyllabified, its phonetic value changes from [l] to [ɾ], as illustrated in *halin* in (1a). Similarly, Korean stop sounds, which are all voiceless, become voiced when placed between two voiced sounds (usually vowels), as in *hankwuke* in (1b) or *patun* in (1d). Vietnamese speakers tend not to apply this voicing rule either if they fail to apply resyllabification, which is caused by their boundary-observing syllable-by-syllable pronunciation.

between read-aloud speech and spontaneous speech. It is hypothesized that the L2 orthographic input in the given scripts would negatively influence the learners' pronunciation, hence more likely to induce learners' errors in resyllabification, and as hypothesized, resyllabification errors were more frequently detected in read-aloud speech than in spontaneous speech. We argue that Vietnamese learners encounter difficulties in applying resyllabification for the following two reasons: (1) unlike Korean phonology, Vietnamese phonology lacks resyllabification; and (2) unlike the Vietnamese orthography (i.e., the Roman alphabet), the Korean orthography discretely marks syllable boundaries through its unique square-block syllable-unit writing convention. Resyllabification errors are a combined product of the contrastive phonological rules and orthographic conventions regarding syllable boundaries between L1 Vietnamese and L2 Korean. More specifically, we propose that Vietnamese L1 speakers demonstrate a higher rate of errors when reading aloud than speaking, particularly due to the mismatch between the boundary-crossing phonology and the boundary-observing orthography of Korean syllables.

2. Resyllabification: Phonologies of L1 Vietnamese and L2 Korean

Resyllabification is a phenomenon where the initial syllabification of words is reshuffled so that the component phones get reassigned to different syllables. It specifically occurs when a syllable with a final consonant (coda) is followed by another syllable with no initial consonant (onset); in this environment, the coda consonant of the preceding syllable comes to occupy the empty onset position of the next syllable, a simple version of which is illustrated in (2). Resyllabification follows the Maximal Onset Principle, i.e., a principle that suggests that consonants that are intervocalic are maximally assigned to the onset (Kahn 1976). This is considered to be a universal principle as the syllable structure of CV is more universal than VC, and is thus applied in many languages so long as it conforms with the language-specific phonotactic and syllable structure constraints.

(2) Resyllabification
V<u>C</u> . V → V . <u>C</u>V

A syllable in Korean allows zero or one consonant for the onset and zero or one consonant for the coda, and as seen in the examples in (1), the coda consonant of the preceding syllable becomes the onset of the following syllable when the next syllable does not have an onset consonant and thus starts with a nucleus vowel. Note that the grapheme ㅇ in the onset position is a mere placeholder in Korean with no phonetic value at all, while it represents the consonant [ŋ] sound in the coda position. This resyllabification can occur within a word, across a morpheme boundary, or even between phrases when spoken fast, as demonstrated in (1). Similarly, English also allows resyllabification, as illustrated in (3), although its syllable structure and phonotactics are quite different from those of Korean.

(3) Resyllabification in English

a. *dreamer*	[ʤri**m**.ɚ]	→	[ʤri.**m**ɚ]	
b. *passed us*	[pɛ·s**t**.əs]	→	[pɛ·s.**t**ʰəs]	(Labov 1997: 155)
c. *sailed over*	[sejl**d**.ovɚ]	→	[sejl.**d**ovɚ]	(Labov 1997: 155)

The resyllabification feature of these languages occasionally becomes a basis for unintentional misunderstandings or intentional puns, as more than one syllabification possibility is available from the same sequence of phones whenever a consonant is placed intervocalically.

(4) a. *mwu**l**.i.ya* (물이야) ‘It's water'	*mwu.**r**i.ya* (무리야) 'It's excessive’
b. *pa**n**.(h)a.na* (반하나) 'attracted?'	*pa.**n**a.na* (바나나) 'banana'
c. *a**n** ice cube*	*a **n**ice cube*
d. *tu**n**e a piano*	*tu**n**a fish*

Vietnamese, on the contrary, does not allow resyllabification (Cho 2006; Tran et al. 2019; Kang et al. 2020; Lee and Kwon 2022). In fact, Vietnamese has a similar syllable structure to that of Korean in that it allows up to one consonant for the onset and up to one consonant for the coda, although the number and the kind of consonants allowed for each position are different (Thompson 1988; Nguyen 1997, 2009; Pham 2009; Kirby 2011; Ngo 2021),[5] as illustrated in Table 1.

5 As a matter of fact, the number of permissible consonants for onset varies between 18 to 21 depending on the dialect or on the view as to how to classify such sounds as glottal stop [ʔ], i.e., whether it is a

Table 1. Syllable structures of Korean and Vietnamese

	Onset (C)	Nucleus (G)V	Coda (C)
Possible syllables	(G)V, C(G)V, (G)VC, C(G)VC		
Korean	18 consonants /p, p', p^h, t, t', t^h, tɕ, tɕ', tɕ^h, k, k', k^h, s, s', h, m, n, l/	10 monophthongs 11 diphthongs	7 consonants /p, t, k, m, n, ŋ, l/
Vietnamese	18 consonants /b, t, d, t^h, c, k, f, v, s, z, χ, ɣ, h, m, n, ɲ, ŋ, l/	11 monophthongs 3 diphthongs	8 consonants /p, t, k, m, n, ŋ, j, w/

In both Korean and Vietnamese, V, GV, CV, CGV, VC, GVC, CVC, and CGVC are possible syllable compositions, and therefore, the resyllabification environment of VC.V is available in Vietnamese too. Nevertheless, Vietnamese L1 speakers tend not to resyllabify the coda consonant in the preceding syllable as the onset of the following syllable when it begins with a vowel; instead, they pronounce each syllable separately as if the syllable boundaries are strictly enforced and should never be crossed. An example of syllable-by-syllable pronunciation of Vietnamese is illustrated in (5a), which contrasts with the resyllabified pronunciations of similar-sounding phrases in English or Korean.

(5) Lack of resyllabification in Vietnamese

a. *Cảm ơn bạn* 'thank you'	[ka:**m**.ə:n.ba:n]	<Vietnamese>
b. *Come on Ben, C'mon Ben*	[kʌ.**m**ən.bɛn]	<English>
c. *kem-un pyen* (검은 변) 'black stool'	[kʌ.**m**un.bjʌn]	<Korean>

Vietnamese is known to be a monosyllabic language in that words mostly consist of a monosyllable or a combination of two or more monosyllabic morphemes. Additionally, Vietnamese is a tonal language, where each syllable has its own tone preassigned. These features seem to make the Vietnamese syllable an unbreakable unit, so resyllabification, which requires crossing the syllable boundary, is impossible in this language. Chinese, which is also a monosyllabic, tonal language, does not allow resyllabification either (Shim 2012), which seems to support this view.

phoneme or an allophone. Note also that the 8 coda consonants include approximants /j/ and /w/, so the Vietnamese coda is alternatively termed G/C.

Another reason for the lack of resyllabification in Vietnamese may be related with a phonetic feature of vowel-initial syllables: that is, there might not exist vowel-initial syllables in the first place, which lack onsets such as V or VC. Some argue that Vietnamese requires an initial consonant in every syllable (Thompson 1988; Pham 2009; Kirby 2011): even those words that start with a vowel graphemically, such as *in* 'print' and *ăn* 'eat', actually have a glottal stop [ʔ] in the onset phonetically (which is not visible in the regular orthography).[6] Therefore, according to this view, the minimal Vietnamese syllable is CV, which would not allow the resyllabification environment of VC.V to be formulated from the beginning; no coda consonant could be intervocalic as no syllable begins with a vowel.

Whatever the reason may be, Vietnamese speakers, whose L1 lacks resyllabification, would obviously have difficulty in applying this rule when they learn and speak Korean as L2 due to a negative transfer. Because of this interference from their L1 phonology, Vietnamese speakers may generate errors in applying the resyllabification rule.[7] However, while this explains why Vietnamese speakers produce resyllabification errors when they learn Korean, it does not account for the observation that they make more errors when reading than when speaking. In the remainder of the paper, we will explore this question in depth.

3. Speech corpus data

3.1 Data collection

This study used the Vietnamese L1 speakers' speech data of Korean from the L2 speakers' speech corpus called "A Collection of Foreigners' Korean Speech Data for AI Training" (Yoon 2021). This corpus consists of speech clips obtained from 1,191 L2 speakers of 80 different nationalities residing in Korea in 2021; the majority of the corpus

6 This view is related with the number of allowable onset consonants, as discussed in footnote 5.

7 Lee and Kwon (2022) studied resyllabification errors using reading tasks performed by 14 native Vietnamese learners and reported a negative phonological transfer from L1 more specifically. They argued that most of the errors were found in the Korean L2 consonants that are not available as onset consonants in their L1 (compare the possible onset consonants between Korean and Vietnamese in Table 1): the rate of applying resyllabification was 86.8% for the consonants that exist as Vietnamese onset, and 62.5% for the non-existing consonants in the Vietnamese onset position.

is concentrated on the data collected from the speakers whose native languages (L1) are Chinese, English, Japanese, Thai, and Vietnamese. The corpus data encompass two speech types: script reading and spontaneous speech. Script reading involves the read-aloud speech of written scripts covering five topic categories related to Korean culture and life; spontaneous speech refers to the participants' responses to questions on similar topics as those covered in the script reading. The script reading consists of 45 scripts covering a range of topics. Each script is composed of 25 sentences, resulting in a total of 1,125 sentences, with an average sentence length of approximately 13 to 15 words. For the spontaneous speech data, participants were presented with 75 questions for each of the five topic categories. They were instructed to provide full-sentence answers to these questions (see Kim et al. (2022) for more detail).

For the current study, we have sampled speech data from a total of 40 Vietnamese speakers: specifically, for each speaker, 75 sentences from the script reading speech and 75 answers from the spontaneous speech are collected respectively. We collected data exclusively from intermediate- and advanced-level speakers, whose TOPIK (Test of Proficiency in Korean) scores ranged from 4 to 6 levels (37 participants—14 for level 4, 13 for level 5, and 10 for level 6) or who were assessed to possess equivalent proficiency (3 participants).[8] We deliberately selected data from non-beginning speakers who are already familiar with Korean pronunciation and orthography and also able to express themselves in full Korean sentences, in order to ensure that their spontaneous speech can be compared with the read-aloud speech. This approach allows us to avoid errors typically associated with beginners who lack sufficient knowledge and experience with the Korean language. All 40 participants are native Vietnamese speakers who have learned Korean for over 5 years and lived in Korea for over 4 years on average, as demonstrated in Table 2.[9]

8 TOPIK (Test of Proficiency in Korean) is a standardized Korean proficiency test for Korean as a second/foreign language conducted by Korean National Institute for International Education. The scores range from Level 1 to 6 with 6 being the highest level. Participants with TOPIK level 4 were assigned to intermediate level and those with TOPIK level 5 and 6 were assigned to advanced level. Three participants without an official score were classified according to their self-assessment scores (one in advanced level and two in intermediate level).

9 Unfortunately, it was hard to find speech data from male speakers who finished both tasks, so we ended up using the female speakers' speech data only. As this is not a pre-designed experiment study, we could not control the conditions, so the imbalance in gender remains as a limitation of this corpus study.

Table 2. Information of selected participants from the L2 speech corpus

L1	Proficiency level	Mean learning period (yrs.)	Mean length of residence in Korea (yrs.)	Mean age (yrs.)	Gender	Education
Vietnamese	Intermediate & advanced	5.5	4.1	30.9	Female	High school or higher

All speech data were transcribed and annotated by native Korean speakers who possessed knowledge in linguistics and successfully passed the pre-training transcription and annotation tests. Subsequently, they underwent multiple training sessions focused on Korean phonology and error annotation. Each sentence is initially annotated by two evaluators and their annotations were cross-checked by three graduate students who made the final decision in case of any discrepancies or mismatches (see Choi et al. (2022) for more information).

3.2 Data overview

The collected spoken data comprises a total of 46,200 *ecel*'s in script reading and 53,606 *ecel*'s in spontaneous speech. An *ecel* (literally a 'word phrase') represents a combination of word stem and grammatical marker(s), separated by a space in Korean writing; for practical purposes, an *ecel* can be approximated as a word in English. As Korean is an agglutinative language, words are normally attached by grammatical markers. As many of these markers start with a vowel such as *i* (이) 'subject marker', *ey* (에) 'in', and *un* (은) 'modifier ending,' many words are consequently situated in the context of resyllabification and do undergo the process. Therefore, we have decided to analyze the realizations of resyllabification based on *ecel* units (we will hereafter refer to an *ecel* as a "word" for the sake of convenience in explanation).

The script reading data and the spontaneous speech data each have 75 individual speech clips per speaker, amounting to a total of 3,000 clips from 40 speakers. The script reading data contains a total of 46,200 words (120,560 syllables) and the spontaneous speech data comprises 53,606 words (146,594 syllables).[10]

10 Even though the numbers of words in terms of tokens are comparable, there exists a noticeable discrepancy in the number of word types between the script reading and the spontaneous speech: there are 814 word types identified in the script reading, while the spontaneous speech consists of 10,515 word types. It is because in the script reading task, all 40 speakers read the identical script comprising 75 sentences, whereas in the spontaneous speech, each speaker provided individualized responses to 75 different questions in terms

Table 3. Composition of speech data

	Speech clips	Words (Token)	Syllables
Script reading	75 (sentences) x 40	46,200	120,560
Spontaneous speech	75 (answers) x 40	53,606	146,594
Total	6,000	99,806	267,154

From the collected data, we extracted and tallied all individual instances of potential resyllabification environments based on the type of coda consonants, as demonstrated in (6).[11] In addition to the typical case illustrated in (6a), we included the consonant clusters in (6b) and the consonants that undergo other phonological rules such as coda neutralization or palatalization. The critical aspect is whether (part of) the coda consonant of the preceding syllable becomes the onset of the next syllable, irrespective of any additional phonetic changes to the consonant.

(6) a. Simple coda: *kwuke* (국어) 'national language' [ku.gʌ];
kongwen-ul (공원을) 'park' [koŋ.wʌ.nɯl]
b. Coda clusters: *kaps-i* (값이) 'price' [kap.si]
c. Coda undergoing neutralization: *kethos* (겉옷) 'coat' [kʌ.dot]
d. Coda undergoing palatalization: *kwuti* (굳이) 'bother to' [ku.dʑi]

On the other hand, three cases were excluded from the calculation of resyllabification environments. The first case is when the coda consonant ㅇ *ng* is followed by a vowel (e.g., *sangtaypangi* (상대방이) [saŋ.dɛ.ba**ŋ**.i] 'counterpart-Nom'). In Korean, [ŋ] is allowed only as a coda and not as an onset. Therefore, the coda ㅇ [ŋ] can never be resyllabified as the onset of the next syllable; as mentioned earlier, the ㅇ in the onset position serves as a dummy placeholder. The second case is when the syllable ends with ㅎ *h* followed by a vowel in the next syllable. In such cases, the ㅎ [h] is deleted (e.g., *coha* (좋아) [tɕo.a] 'like'). The third case involves the coda consonant ㄴ [n] followed by [i] or [j], resulting in the duplication of ㄴ, occupying both the coda and onset positions (e.g., *sinhonyehayng* (신혼여행) [sin.ho**n.nj**ʌ.hɛŋ] 'honeymoon'). This case is excluded from resyllabification because of the reason that the coda stays in its

of content and length. This variability in answers leads to a much broader range of vocabulary usage, thus resulting in a significantly higher number of word types.

11 We focused solely on the resyllabification within words (*ecel*'s) as resyllabification between phrases across word (*ecel*) boundaries is optional and depends on factors such as speech rate and presence of pauses.

original position.

Table 4 presents the frequencies of the words necessitating resyllabification. Both types of speech exhibit a similar rate of instances of resyllabification environment. In the script reading data, there are a total of 10,760 word tokens with resyllabification environment (23.3% of the total words), and the spontaneous speech data contains a total of 12,386 instances (23.1%).

Table 4. Potential words that require resyllabification

	Resyllabifiable words (token)	Percentage per total words (%)
Script reading	10,760	23.3
Spontaneous speech	12,386	23.1
Total	23,146	23.2

4. Data analysis and results

4.1 Resyllabification errors: Script reading vs spontaneous speech

In the sampled speech data taken from 40 Vietnamese speakers from the Korean L2 speakers' speech corpus, 1,142 instances of resyllabification errors were identified in the script reading speech, whereas 988 instances of such errors were yielded in the spontaneous answer speech. These results, detailed in Table 5, indicate a 10.61% error rate for script reading and 7.98% for spontaneous response speech. A Chi-Square test further verifies that the script reading speech exhibits a significantly higher rate of resyllabification errors than the spontaneous speech [χ^2 (1, N = 23,146) = 47.906, $p < .00001$].

Table 5. Error rates in script reading and spontaneous speech

	Script reading	Spontaneous speech	Total
Raw error frequency	1,142	988	2,130
Raw total frequency	10,760	12,386	23,146
Error rate (%)	10.61	7.98	9.20

The fluency level of speakers does not appear to influence the result. As demonstrated in Table 6, Vietnamese speakers consistently exhibit a significantly higher rate of errors

in script reading compared with spontaneous speech irrespective of their fluency level, which is measured by their TOPIK scores.

Table 6. Error rates per fluency level in script reading and spontaneous speech

TOPIK level	N of speakers	Script reading			Spontaneous speech			p-value
		Errors	Total	Error rate (%)	Errors	Total	Error rate (%)	
4	14	370	3,752	9.86	317	3,866	8.20	.0113
5	13	393	3,484	11.28	370	4,267	8.67	.0001
6	10	278	2,680	10.37	244	3,577	6.82	<.00001

As hypothesized, the results confirm that Vietnamese learners of Korean do produce more resyllabification errors when reading scripts than speaking spontaneously. We have already predicted in Section 2 that Vietnamese L1 speakers would have difficulty applying resyllabification in their Korean speech because of the negative transfer from their L1 phonology. However, given that phonology would influence their pronunciation regardless of the speech type, it does not account for the discrepancy in the error rate between the script reading speech and the spontaneous speech. This suggests that there may be factors other than the negative phonological transfer contributing to this outcome, and three potential explanations arise. First, the reading scripts and the spontaneous speech might have different levels of difficulty in terms of vocabulary and/or grammar (such as the use of grammatical markers); namely, the words and phrases in the written scripts might be more advanced and formal than those used in the spontaneous speech and thus could lead to an increased error rate. Second, during the script reading task, speakers might come across unfamiliar words they have never used before, whereas in the spontaneous speech, they are likely to use the vocabulary they are well-acquainted with from everyday use, possibly leading to fewer errors. Lastly, orthographic influence could contribute to an increased error rate as it could negatively affect learners' production of L2 speech sounds. We will explore the first two potential explanations in the following subsection.

4.2 Difficulty of vocabulary

To explore the causes of different resyllabification error rates between the read-aloud

speech and the spontaneous speech, we first hypothesized that the level of words in the script reading speech might be more difficult than that in the spontaneous speech. To test this, we classified all the words that appeared in each data set by levels of difficulty based on the guidelines proposed by the "Applied Research for the International Standard Curriculum of Korean Language" (Kim 2017) and also on the vocabulary lists from a set of Korean language textbooks (i.e., *Ewha Korean* vols. 1-6). Tables 7 and 8 show the tokens of the content word stems and the grammatical endings respectively, classified by three difficulty levels.[12]

Table 7. Difficulty levels of content word stems

		Beginning	Intermediate	Advanced	Total
Script reading	Raw frequency	8,520	1,960	80	10,560
	Percentage (%)	80.68	18.56	0.76	100.00
Spontaneous speech	Raw frequency	10,334	1,246	162	11,742
	Percentage (%)	88.01	10.61	1.38	100.00

Table 8. Difficulty levels of grammatical (functional) markers

		Beginning	Intermediate	Advanced	Total
Script reading	Raw frequency	9,600	600	40	10,240
	Percentage (%)	93.75	5.86	0.39	100.00
Spontaneous speech	Raw frequency	10,449	599	161	11,209
	Percentage (%)	93.22	5.34	1.44	100.00

Contrary to our initial assumption, determining the level of difficulty between the two types of speech is not straightforward. In both the scripts and the spontaneous responses, the majority of the utilized words and grammatical markers are of a beginning level (over 80% for content word stems; over 90% for functional markers), followed by an intermediate level, with the least used being of an advanced level. However, it is hard to judge which type of speech used more difficult levels of vocabulary. As for the content word stems, interestingly, the spontaneous speech used a significantly higher rate of the beginning-level [χ^2 (1, N = 22,302) = 228.3684, p < .00001] and also of the

12 The total number of word stems and grammatical endings shown in Tables 7 and 8 are different from the total frequency of resyllabifiable words introduced in Table 4 because there are words without grammatical endings (e.g., *halin* 할인 'discount') or without word stems (e.g., *ttaymwuney* 때문에 'because') in the data sets.

advanced-level vocabulary [$\chi^2(1, N = 22,302) = 20.0462, p < .00001$], while the script reading used a higher rate of the intermediate-level one [$\chi^2 (1, N = 22,302) = 285.4188, p < .00001$]. As for the grammatical endings, only the advanced-level vocabulary demonstrates a difference of statistical significance [$\chi^2 (1, N = 21,449) = 63.0364, p < .00001$]; in this category, the spontaneous speech indicates a higher frequency. In other words, all in all, it is simply hard to determine which type of speech presents a more difficult level of vocabulary. As the speakers are intermediate and advanced learners of Korean, they seem to be able to make use of beginning to advanced vocabulary in their spontaneous speech on par with that of the scripts for both content and function. Therefore, we should rule out the level of vocabulary as a factor for the higher rate of resyllabification errors in the script reading speech.

4.3 Familiarity of vocabulary

The second potential factor to consider is familiarity of vocabulary. Even for the vocabulary of the same level, the words utilized in spontaneous speech, compared with those used in the scripts, would generally be more familiar to the speakers, as these are the words they hear and use frequently in everyday life. This familiarity would allow them to pronounce these words "correctly," employing the appropriate resyllabified pronunciation. Under the assumption that the words used in spontaneous speech are "familiar," we would expect the equivalent rate of resyllabification when these same words appear in the scripts as well. To investigate this, we sorted out the identical words used in both the script reading and the spontaneous responses to compare their respective error rates. In total, we found 66 resyllabifiable words that were common to both data sets. However, most of these overlapping words showed a significant disparity in frequencies between the two types of speech. Therefore, we only included those words that had relatively minor differences in frequency (less than double the gap). We also excluded those words with only a couple of total errors. As a result, 13 out of 66 words were selected based on their total occurrences in the data. The list of words with comparable total frequencies is illustrated in Table 9.

Table 9. Error rates of the overlapping words

Overlapping words	Script reading			Spontaneous answer		
	Error (freq.)	Total (freq.)	Error rate (%)	Error (freq.)	Total (freq.)	Error rate (%)
ttaymwuney 때문에 'because'	23	200	11.50	11	147	7.48
manh-un 많은 'many'	5	40	12.50	1	55	1.82
manh-i 많이 'a lot'	24	360	6.67	17	364	4.67
mek-ul 먹을 'eat-Mod(Fut)'	16	40	40.00	6	37	16.22
mom-i 몸이 'body-Nom'	2	40	5.00	0	21	0.00
pat-un 받은 'receive-Mod(Pst)'	12	40	30.00	10	51	19.61
pyengwen-ey 병원에 'to a clinic'	8	80	10.00	1	46	2.17
sikan-i 시간이 'time-Nom'	3	40	7.50	3	54	5.56
umsik-ul 음식을 'food-Acc'	4	40	10.00	5	47	10.64
umsik-i 음식이 'food-Nom'	13	40	32.50	10	50	20.00
cen-ey 전에 'before'	15	120	12.50	7	101	6.93
cip-ey 집에 'at home'	7	40	17.50	5	53	9.43
choykun-ey 최근에 'recently'	8	40	20.00	9	66	13.64

Surprisingly, the result reveals that even the identical words exhibit higher error rates in script reading than in spontaneous speech: out of the 13 selected words, 12 displayed a higher resyllabification error rate in script reading and one remaining word (i.e., *umsik-ul* 'food-Acc') exhibited a comparable rate in both contexts. On average, the error rate of these 12 words is 17.14% in the script reading, nearly double the 8.96% rate found in spontaneous responses [χ^2 (1, N = 2,212) = 13.4591, p < .000244].

More dramatically, the same individual speaker who applied resyllabification during spontaneous speech did not resyllabify when reading the identical word from a script. Table 10 demonstrates a few such examples.

Table 10. Resyllabification of the overlapping words by the same speakers

Data / Speaker	Korean word	Script reading	Spontaneous speech
S1	*ton-i* 돈이 'money-Nom'	[to**n**.i]	[to.**n**i]
S2	*choykun-ey* 최근에 'recently'	[t͡ɕwe.gu**n**.e]	[t͡ɕwe.gɯ.**n**e]
S3	*hankwuk-eyse* 한국에서 'in Korea'	[han.gu**k**.e.sʰʌ]	[han.gu.**g**e.sʌ]

These examples strongly suggest that neither the level of difficulty of vocabulary nor the speaker's familiarity with them significantly contributed to the higher rate of resyllabification errors in the script reading speech. This leads us to the final hypothesis:

orthography may be the primary influence on the resyllabification errors by native Vietnamese learners of Korean. This will be explored in the following section.

5. Discussion

As noted earlier in Section 2, the phonological syllable boundary of Vietnamese is more rigid than that of Korean in that Vietnamese does not allow resyllabification while Korean does. It is therefore understandable that Vietnamese speakers make errors in resyllabification when they speak Korean as L2 because of a negative transfer from their L1 phonology. Yet, as observed in Section 4, their performances in resyllabification differ between read-aloud speech and spontaneous speech: namely, not only the overall error rate but also the error rate of the identical list of words produced by the identical group of speakers is significantly higher in the read-aloud speech than in the spontaneous speech, even if the levels of vocabulary used in the two data sets are not very different from each other. This indicates that the L2 orthography of Korean must have influenced the speakers' read-aloud speech negatively.

5.1 Orthographic influence on L2 speech production

Korean uses its own writing system called Hangeul, which is a transparent phonographic orthography. An orthography is deemed to be transparent if the phoneme-to-grapheme correspondence is consistent. In most cases, one sound corresponds to a single letter or grapheme in Korean, but there are also phonological rules that make it deviate from the initial transparent mapping between phonemes and graphemes. Vietnamese writing system, on the other hand, is based on the Roman alphabet, which was first invented by a French missionary Alexandre de Rhodes, and it substituted the Vietnamese first writing system, Chữ Nôm, derived from the Chinese characters (Thompson 1988). Unlike the former Chinese-like writing system, the new Vietnamese alphabet has a clear grapheme-phoneme correspondence, which also makes it a transparent system.[13]

13 Vietnamese does not normally have phonological variations because each syllable has its original pronunciation. As discussed later, the phonological syllable boundary of Vietnamese is rigid just as in Chinese so that there are few phonological changes between phones in contact. If a Vietnamese word does show

Orthography may exert a positive or negative influence on learners' pronunciation of the target language. Contrary to expectations, an L2 orthography that differs from that of L1 does not necessarily result in a negative effect. For example, Thai learners of Korean, whose L1 orthography is different from L2 Korean, have been reported to achieve high scores of fluency, accuracy, and intonation in pronunciation if they had a greater orthographic awareness of Korean final consonants (Kim 2022). Similarly, Chinese learners' exposure to the letter ㅎ *h* in Korean has been found to enhance their production of ㅎ /h/ allophones (Han and Kim 2017).

On the other hand, an L2 orthography that is similar to that of L1 does not necessarily lead to a positive effect either. For instance, Bassetti (2006) discovered that Pinyin—the Roman-alphabet-based orthography of Chinese devised to facilitate pronunciation—actually hindered the beginning learners who were native readers of the Roman alphabet from yielding target-like pronunciations. Similarly, in Rafat's (2011) study, English L1 learners of Spanish produced pronunciation errors caused by L1 transfer induced by the orthographic input from L2 Spanish: they mispronounced silent *h* as [h] ([haɾapo] for *harapo* [aɾapo]), [z] instead of *z* [s] ([zatara] instead of *zatara* [satara]), [l] instead of *l* [j] ([polero] instead of *pollero* [pojeɾo]), etc. Young-Scholten and Langer (2015) have also documented that the orthographic input in L2 German interfered with native English-speaking learners' target-like speech production, particularly with the letter *s*, which represents [z] in German but [s] in English (sie 'she' [zi:] pronounced as [si:]).

However, it is evident that a distinct L2 orthography can potentially create confusion in learners' acquisition of L2 phonology, and the Korean orthography serves as an example. Studies have shown that the readers of the Roman alphabet are negatively affected by the Korean L2 orthography. For instance, French speakers learning Korean have been reported to mistakenly replace ㅇ /ø/ with ㅎ /h/ (미안해요 /mianhɛjo/ 'sorry' as [mi**h**anhɛjo]), confused by the visual similarity between the two letters. Yim (2007) argues the mispronunciation is caused by L2 orthography as well as L1 interference of French uvular [ʁ]. Another study demonstrates that the Korean orthography has a

---

a phonological variation, it is already reflected in standard spelling (Cho 2005). For example, Vietnamese number 15 *mười lăm* is originally a combination of 10 *mười* and 5 *năm*, but the spelling *lăm* already reflects the changed pronunciation of [n] as liquid [l], which makes the phoneme-grapheme correspondence in Vietnamese more transparent. It is comparable to 달걀 *tal.g(y)al* 'egg' in Korean, for instance, which already orthographically reflects the phonological change derived from 닭알 *talk.al* 'chicken's egg.'

negative impact on American learners' pronunciation of Korean liquid ㄹ *l*, which has two phonetic realizations [l] or [ɾ] depending on the position. According to Kwon and Park (1999), the participants consistently mapped Korean ㄹ *l* to English /l/ regardless of its position, failing to properly pronounce the allophones. These finds emphasize the potential challenges posed by L2 orthographic systems, as they can hinder learners' accurate acquisition of L2 phonology.

In addition to these errors in grapheme-to-phoneme mapping, orthography has actually been found to influence resyllabification errors. In her research solely focusing on orthographic influences on the phonological development of Korean, Lee (2012) discovered that American English speakers learning Korean made more pronunciation errors with intervocalic obstruents when they appeared in a coda position followed by a vowel-initial syllable (집이 /cip.i/ 'house-Nom') compared to when they were in the onset position (노비 /no.pi/ 'slave'), when asked to read these words in paired conditions. In other words, American learners demonstrated higher accuracy in words without a resyllabification condition than those requiring resyllabification.

As discussed in Section 2, English is a language that allows for resyllabification: in English, an intervocalic consonant is often resyllabified as the onset of the following syllable. However, Lee's (2012) study suggests that even speakers of L1 where phonological syllable boundaries are fluid enough to permit resyllabification may encounter difficulties in applying resyllabification in an L2. In the case of English learners struggling with resyllabification in Korean, it is important to note that this difficulty cannot be solely attributed to a negative transfer from L1 phonology; instead, it must be more closely related to orthography. If English L1 learners have difficulties with resyllabification in Korean, it is not surprising to find Vietnamese speakers struggling with similar issues. This is because Vietnamese uses the Roman alphabet like English, but its L1 phonology does not allow for resyllabification.

5.2 Orthography with rigid syllable boundary

Korean orthography Hangeul stands out with its unique visual configuration as a phonographic writing system. Languages including English and Vietnamese that adopt a phonographic writing system like the Roman alphabet share the general orthographic principle of displaying individual graphemes in a left-to-right linear sequence to represent

the sequence of phonemes in the same linear order. However, Korean letters are not presented linearly but rather in square blocks containing two or three letters each. For example, the word *hankwuke* 'Korean' consists of 8 letters in the Korean alphabet Hangeul and if displayed linearly, would look like ㅎ ㅏ ㄴ ㄱ ㅜ ㄱ ㅇ ㅓ. But instead, it is written in 3 syllable blocks of letters, looking like 한국어, as illustrated in Figure 1. Each syllable occupies a single character space (Simpson and Kang 2004: 139).

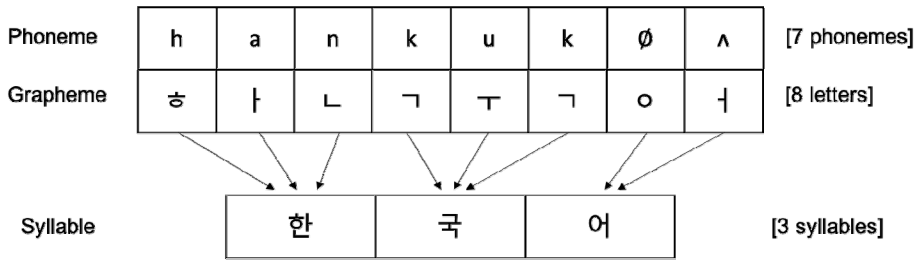

Figure 1. Orthographic convention of syllable blocks in Korean

As such, the Korean orthography employs a graphical representation where a syllable is visually depicted as a square block. Compared to other phonographic writing systems, Hangeul is visually more compositional as it forms a visibly distinct intermediate unit of a syllable. In languages that use the Roman alphabet, it is difficult to visually differentiate one syllable from the adjacent syllable within a word and thus the orthographic syllable boundary can be ambiguous, as is the phonological syllable boundary, as discussed in Section 2. In contrast, Korean Hangeul visually packages letters into syllable blocks, providing clear demarcation between neighboring syllables. As a result, the orthographic syllable boundary in Korean is always visually firm and never ambiguous.

In this aspect of the syllable-block presentation, Hangeul differs from other phonographic writing systems and bears a resemblance to morphographic orthographies like Chinese Hanzi (or Chinese-character-based Japanese Kanji). In Hanzi, each character, shaped in a square block, represents a morpheme, which is also a syllable; thus, each character inherently represents a syllable, and the boundaries between characters coincide with morpheme boundaries, resulting in rigid syllable boundaries. Noting this syllabic characteristic of Hangeul writing, Taylor (1980) argues that Korean writing system is an "alphabetic syllabary," combining features of an alphabet, a syllabary, and a logography.

Similarly, Pae (2018) suggests that Hangeul can be described as a phonemic-syllabary or "alphasyllabary," as it exhibits dual characteristics of both a phonemic writing system and a syllabic writing system.

Resyllabification involves redefining the syllable boundary by shifting the boundary before the preceding coda consonant as it becomes a new onset. This process is possible due to the fluid nature of the phonological syllable boundary in Korean. In contrast, the orthographic syllable boundary in Korean, as depicted in Figure 1, remains rigid due to the syllable-block writing convention of Hangeul, which is similar to Chinese Hanzi.[14]

Table 11. Rigidity of syllable boundaries in phonology and orthography

	Chinese	**English**	**Korean**	**Vietnamese**
Phonology	rigid	fluid	fluid	rigid
Orthography	rigid	fluid	rigid	fluid

Due to the disparity between the phonological fluidity and the orthographic rigidity of the Korean syllable boundary, L2 learners of Korean often experience difficulties in perceiving and producing syllables. This is evident in the case of American English speakers who successfully apply resyllabification in L1 struggle to do so in L2 Korean during reading tasks (Lee 2012). The syllable-block orthography provides the participants with misleading visual cues to retain the original syllable boundaries, preventing the application of resyllabification. We argue that this is exactly why Vietnamese learners who are proficient enough to apply resyllabification in spontaneous speech despite their rigid L1 phonology fail to apply it in script reading. As a result, they produce more errors in read-aloud speech compared to spontaneous speech. The rigid syllable boundary in Korean orthography offers visual signals that discourage resyllabification, thus negatively influencing their L2 pronunciation.[15]

14 Similar to the orthographic syllable boundary, the phonological syllable boundary in Chinese is also rigid, lacking the ability for resyllabification (Shim 2012). Consequently, it is not surprising that Chinese learners of Korean have been reported to pronounce words at the syllable-unit level, often struggling with applying the phonological rules of Korean that involve phonetic contacts and changes at syllable boundaries (see Kim (2018) for a case of nasalization).

15 Another potential orthographic effect affecting resyllabification may be speculated. It is the presence of the dummy placeholder ㅇ in the onset position of the vowel-initial syllables. As mentioned earlier, the letter ㅇ represents consonant [ŋ] in the coda position. Although Korean phonology does not allow [ŋ] as a permissible onset consonant, Vietnamese does (see Table 1 again for comparison). Therefore, it is plausible that Vietnamese learners perceive the letter ㅇ in the onset position as a substantial consonant and phonetically realize the onset ㅇ as [ŋ], similar to their native language. Then, an ㅇ-initial syllable may

6. Conclusion

This study has examined the resyllabification errors produced by Vietnamese learners of Korean, utilizing an L2 speech corpus composed of script reading and spontaneous speech, and discovered that the error rate is higher in script reading than in spontaneous speech. Among the three hypotheses proposed to explain the disparity in error rates between read-aloud and spontaneous speech, we conclude that the L2 orthographic input in the read-aloud speech contributes to the higher rate of resyllabification errors and that the alternative hypotheses such as the differences in the level of vocabulary or in the learners' familiarity with the words between the two speech types have been found to be invalid.

We argue that the Vietnamese L1 phonology, which enforces a rigid syllable boundary, contrasts with the fluid syllable boundary in L2 Korean and that this difference in phonology serves as a negative transfer, leading Vietnamese learners to experience difficulties in resyllabification when speaking Korean. Moreover, the syllable-block writing convention of the Korean orthography presents a visually rigid syllable boundary, whereas the Vietnamese L1 orthography (i.e., the Roman alphabet) does not impose such rigidity. Consequently, the orthographic rigidity of syllable boundary of Korean provides Vietnamese learners with negative visual signals so that in script reading, where the orthographic input is provided, the speakers tend to produce more errors than in the spontaneous speech with no such input. In summary, the Korean L2 orthography exerts a negative effect on the speech performance of Vietnamese learners in terms of resyllabification.

Using Korean Hangeul as a case, the current study has explored the influence of L2 orthography on learners' phonology and demonstrated that not only the individual graphemes but also their orthographic syllable configurations can affect their phonetic realizations. Specifically, the current analysis—based on the discrepancy between the phonological and orthographic fluidity/rigidity of syllables—offers insights into predicting the likelihood of resyllabification errors depending on the syllable type of the first language (L1). We anticipate that future studies involving speakers from all four language types (sorted by rigidity level) will provide a deeper understanding about the role of

---

not be perceived as vowel-initial, thereby not constituting a resyllabification environment. While this hypothesis cannot be definitely confirmed in this non-experimental study, it is worth noting that we did not observe any such cases of [ŋ] pronunciation of ㅇ in the onset position in our speech dataset.

rigidity of phonological and orthographic syllable boundaries. However, this study has its limitations as a corpus study using large-scale data annotated for AI training: while providing an extensive amount of naturally spoken data, it lacks rigorous control over variables that a pre-designed experiment study would have. We hope these limitations will be addressed in subsequent research.

References

Bassetti, Benedetta. 2006. Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language and Literacy* 9(1): 95-114. https://doi.org/10.1075/wll.9.1.07bas

Bassetti, Benedetta. 2008. Orthographic input and second language phonology. In Thorsten Piske and Martha Young-Scholten (eds.), *Input matters in SLA*. Clevedon, UK: Multilingual Matters.

Cerni, Tania, Bene Bassetti, and Jackie Masterson. 2019. Effects of orthographic forms on the acquisition of novel spoken words in a second language. *Frontiers in Communication* 4(31): 1-14. https://doi.org/10.3389/fcomm.2019.00031

Cho, Myeong Sook. 2005. A study on Korean grammar teaching in Vietnam universities and contrastive linguistic analysis of Korean with Vietnamese language [in Korean]. *Journal of Korean Language Education* 16: 321-350.

Cho, Myeong Sook. 2006. A contra-linguistic study of Korean and Vietnamese [in Korean]. *Proceedings of the 16th International Conference on Korean Language Education*, 363-391.

Choi, Hye-Won, Soo-Yeon Kim, and Gunhee Ko. 2022. An analysis of segmental accuracy of Korean for automated assessment: With focus on Chinese, Vietnamese, and Thai speakers [in Korean]. *Journal of the International Network for Korean Language and Culture* 19(1): 221-247.

Erdener, V. Doğu and Denis K. Burnham. 2005. The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning* 55(2): 191-228. https://doi.org/10.1111/j.0023-8333.2005.00303.x

Escudero, Paola, Ellen Simon, and Karen E. Mulak. 2013. Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition* 17(2): 384-395. https://doi.org/10.1017/S1366728913000436

Han, Jeong-Im and Joo-Yeon Kim. 2017. The influence of orthography on the production of alphabetic, second-language allophones by speakers of a non-alphabetic language. *Journal of Psycholinguistic Research* 46(4): 963-982. https://doi.org/10.1007/s10936-016-9474-7

Hwang, Mi-kyeong. 2016. Analysis of Korean liquid pronunciation of Vietnamese entry-level students [in Korean]. *Ewha Language and Literature* 39: 49-87.

Jang, Hye-jin. 2018a. Acoustic analysis of Korean word-initial stops in Vietnamese learners [in

Korean]. *The Studies of Korean Language and Literature* 62: 375-400. http://doi.org/10.15711/06212

Jang, Hye-jin. 2018b. Analysis of pronunciation types and acoustic properties in Korean word-initial affricate of Vietnamese Korean learners [in Korean]. *Journal of Korean National Language and Literature* 61: 363-385.

Kahn, Daniel. 1976. *Syllable-based generalizations in English phonology*. PhD Dissertation. Massachusetts Institute of Technology.

Kang, Eun-jin, Thi-houng Tran, and Chae-hyung Cho. 2020. A study on the Vietnamese learner's Korean intonation phonetic research-focused on university students learning the Korean language as a foreign language in Vietnam [in Korean]. *The Journal of Korean Language and Literature* 36: 191-219. http://doi.org/10.24227/jkll.2020.02.36.191

Kim, Chun-Ye. 2022. The remarkable effect of Korean orthographic awareness on Thai KFL learners' pronunciation production. *Linguistic Research* 39(Special Edition): 123-158. https://doi.org/10.17250/khisli.39..202209.005

Kim, Hyunjin, Inkyung Lee, Yeorim Kang, Eunsun Park, and Miyoung Kim. 2010. *Ewha Korean 2-1* [in Korean]. Seoul: Epress.

Kim, Jin-ho. 2018. A study on the teaching method of Korean pronunciation education for Chinese learners: Mainly with nasalized error phenomenon [in Korean]. *The Journal of Asian Cultural Studies* 47: 123-142.

Kim, Jung-sup. 2017. *Applied research for the international standard curriculum of Korean language* [in Korean] (Report No. 11-1371028-000699-01). National Institute of Korean Language. https://www.korean.go.kr/front/reportData/reportDataView.do?mn_id=45&report_seq=932&pageIndex=1

Kim, Soo-Yeon, Hyomin Min, and Hye-Won Choi. 2022. A strategic design and construction of non-native voice data set of Korean speech for AI model training [in Korean]. *The Journal of Linguistic Science* 100: 63-88. http://doi.org/10.21296/jls.2022.3.100.63

Kim, Yeonjin, Inkyeong Lee, Yeorim Kang, Eunseon Park, and Miyoung Kim. 2010. *Ewha Korean 2-2* [in Korean]. Seoul: Epress.

Kim, Young-Seon. 2004. A study on the teaching method of Korean glottalization in Vietnamese students [in Korean]. *Journal of Korean Language Education* 15(2): 51-73.

Kirby, James P. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41(3): 381-392. https://doi.org/10.1017/S0025100311000181

Koo, Jaehee, Jinhee Hyun, Soyong Lee, Sunyoung Hwang, and Minsun Kim. 2011. *Ewha Korean 3-1* [in Korean]. Seoul: Epress.

Koo, Jaehee, Jinhee Hyun, Soyong Lee, Sunyoung Hwang, and Minsun Kim. 2011. *Ewha Korean 3-2* [in Korean]. Seoul: Epress.

Kwon, Pyong-ro and See-gyoon Park. 1999. Some problems of English speakers learning Korean in pronouncing Korean liquids and double consonants [in Korean]. *Korean Language Education* 100: 153-171.

Labov, William. 1997. Resyllabification. In Frans L. Hinskens, Roeland van Hout, and W. Leo Wetzels (eds.), *Variation, change, and phonological theory*, 145-179. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Lee, Duyong. 2019. The production of Korean stops by Vietnamese beginner learners [in Korean]. *Korean Linguistics* 82: 73-94. https://doi.org/10.20405/kl.2019.02.82.73

Lee, Hyun-Hee and Kyung-Keun Kwon. 2022. A study on Korean liaison in prosodic units of Vietnamese learners [in Korean]. *Korean Journal of Linguistics* 25(3): 383-403. http://doi.org/10.35902/URM.2022.68.265

Lee, Jungyeon, Minkyeong Lee, Minjeong Kim, Sooyeon Park, and Soonmi Song. 2012. *Ewha Korean 5* [in Korean]. Seoul: Epress.

Lee, Mihye, Hyunjin Kim, Jaehee Koo, Young Yoon, Soohaeng Lee, Kyeongmi Kwon, and Sunhee Yoo. 2010. *Ewha Korean 1-1* [in Korean]. Seoul: Epress.

Lee, Seonmi and Namhyun Cho. 2022. A study on the correlation between the acquisition pattern of Korean phonological variation and pronunciation learning strategies of learners in Northern Vietnam [in Korean]. *Journal of Education & Culture* 28(6): 251-271. https://doi.org/10.24159/joec.2022.28.6.251

Lee, Soohaeng, Jungmin Cho, Jincheol Park, Insook Cho, and Sooyeon Park. 2012. *Ewha Korean 6* [in Korean]. Seoul: Epress.

Lee, Sooyeon. 2012. *Orthographic influence on the phonological development of L2 learners of Korean*. PhD Dissertation. The University of Wisconsin-Milwaukee.

Ministry of Justice. 2021. *Number of foreigners living in Korea by nationality and age* [in Korean]. Ministry of Justice. https://kosis.kr/statHtml/statHtml.do?orgId=111&tblId=DT_1B040A6&conn_path=I2

Ngo, Binh. 2021. *Vietnamese: An essential grammar*. London; New York: Routledge.

Nguyen, Dình-Hoà. 1997. *Vietnamese*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Nguyen, Dình-Hoà. 2009. Vietnamese. In Bernard Comrie (ed.), *The world's major languages*, 777-796. New York: Oxford University Press.

Pae, Hye K. 2018. The Korean writing system, Hangul, and word processing. In Hye K. Pae (ed.), *Writing systems, reading processes, and cross-linguistic influences* (Bilingual Processing and Acquisition 7), 335-352. Amsterdam: John Benjamins Publishing Company. http://doi.org/10.1075/bpa.7.16pae

Pham, Andrea Hoa. 2009. The identity of non-identified sounds: Glottal stop, prevocalic /w/ and triphthongs in Vietnamese. *Toronto Working Paper in Linguistics* (TWPL) 34: 1-17.

Rafat, Yasaman. 2011. *Orthography-induced transfer in the production of novice adult English-speaking learners of Spanish*. PhD Dissertation. University of Toronto.

Shim, Minhee. 2012. Error analysis on the liaison of Chinese learners based on optimality theory [in Korean]. *The Korean Language and Literature* 115: 55-77.

Simpson, Greg B. and Hyewon Kang. 2004. Syllable processing in alphabetic Korean. *Reading and*

Writing 17: 137-151. http://doi.org/10.1023/B:READ.0000013808.65933.a1

Taylor, Insup. 1980. The Korean writing system: An alphabet? a syllabary? a logography? In Paul A. Kolers, Merald E. Wrolstad, and Herman Bouma (eds.), *Processing of visible language* (NATOCS 13), 67-82. Boston: Springer. https://doi.org/10.1007/978-1-4684-1068-6_5

Thompson, Laurence C. 1988. *A Vietnamese reference grammar*. Honolulu: University of Hawaii Press.

TOPIK. 2021. *Statistics of applicants, examinees, and passers from 2017-2021 by nation and test* [in Korean]. TOPIK. https://www.topik.go.kr/TWGUID/TWGUID0010.do

Tran, Thi Thuy Hien, Nathalie Valleé, and Lionel Granjon. 2019. Effects of word position on the acoustic realization of Vietnamese final consonants. *Phonetics* 76(1): 1-30. https://doi.org/10.1159/000485103

Yim, Seong-Sook. 2007. Error pattern and interference features of French-speaking learners of KSL/KFL. *Journal of Language Sciences* 14(1): 159-181.

Yoo, Soyoung and Hyonhwa Kang. 2018. Analysis of phonemic errors of Korean learners according to language and proficiency [in Korean]. *Language Facts and Perspectives* 44: 357-397.

Yoon, Hee-woo. 2021. *A Collection of Foreigners' Korean Speech Data for AI Training* [in Korean]. CSLEE. https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=505

Yoon, Young, Yeon-im Heo, Kyeongmi Kwon, Eulji Lee, and Woojeon Choi. 2011. *Ewha Korean 4* [in Korean]. Seoul: Epress.

Young-Scholten, Martha and Monika Langer. 2015. The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics* 36(1): 93-114. https://doi.org/10.1017/S0142716414000447

Appendix

Sentences for script-reading speech

1. 벌써 이번 달 월급을 반 이상 써 버렸어. 월급 받은 지 일주일(1주일)밖에 안 됐는데, 대체 어디에다 돈을 쓴 걸까?
2. 내가 뭘 샀는지 기억조차 안 나네. 지금 입고 있는 치마도 처음 보는 것 같은데, 그것도 새로 산 것 같아.
3. 응, 치마도 이번에 새로 산 건데, 정말 예쁘지 않아? 우리 언니랑 지난주 일요일에, 백화점에 갔다가 샀어.
4. 다리도 날씬해 보이고, 키도 커 보이는 게, 나한테 아주 잘 어울려. 꽤 비싸 보이는데, 얼마 주고 샀는지 알아?
5. 한 벌에 이십오만 원(250,000원)인데, 디자인이 너무 마음에 들어서, 빨간색이랑, 주황색, 분홍색까지 다 사 버렸어.
6. 이제 보니, 내가 돈이 부족한 이유가 다 있었네. 내 월급에 비해서, 지출이 너무 많은 것 같아.
7. 한 벌만 사도 충분한데, 내가 또 과소비를 한 것 같아서 지금 생각해 보니 조금 후회가 많이 돼.
8. 한 벌만 입고, 나머지 두 벌은 환불하면 되겠다. 아무래도 오늘 퇴근하면서 백화점에 들러서 환불해야겠어.
9. 너는 평소에 저축은 좀 하고 있니? 나도 정말 그러고 싶은데, 생활비가 빠듯해서 돈을 모을 여유가 없어.
10. 저축이 얼마나 중요한데, 전혀 안 하고 있다니, 어떻게 하려고 그래? 그럼 앞으로 뭘, 어떻게 해야 할까?
11. 우선 저축할 돈과 사용할 돈을 구분해야 해. 그렇지 않으면, 눈 깜짝할 사이에 다 써 버리고 말 테니까.
12. 미리 돈을 나눠서 저금하라는 거지? 그리고 생활비도 월세, 식비, 통신비 등으로 나눠서 예산을 짜야 해.

13. 그렇구나! 그러면 지금 받는 월급에서 삼분의 일(1/3) 정도는 저금하고, 나머지만 사용하는 게 좋겠어.
14. 그래, 그거야. 그리고 불필요한 지출을 줄이면서, 낭비하지 않는 습관을 들이면, 너도 저축할 수 있을 거야.
15. 너는 주로 어디에 돈을 많이 써? 나는 유행에 좀 민감한 편이라, 옷하고 신발을 사는 데 돈을 많이 쓰는 것 같아.
16. 계절이 바뀔 때마다 쇼핑하는 건 물론이고, 유명한 연예인들이 방송에 입고 나온 옷을 사기도 해.
17. 그럼, 네가 좋아하는 블랙핑크가 입었던 옷도 산 적 있어? 나는 지수가 전에 입고 나왔던 겉옷을 산 적이 있어.
18. 무슨 브랜드 옷인지 어떻게 알았냐면 인터넷으로 검색해서 똑같은 옷을 파는 쇼핑몰을 찾았지.
19. 그럼 인터넷으로 주문한 거야? 내가 직접 주문할 때는, 바로 품절돼서, 직접 그 가게까지 찾아가서 샀어.
20. 마침 내가 갔을 때 십 퍼센트(10퍼센트) 할인 행사가 있어서, 다행히 인터넷보다 좀 더 저렴한 가격으로 살 수 있었어.
21. 세일하는데, 그것 말고 다른 건 안 샀어? 나는 출근할 때 입어도 될 만한 바지가 있길래 샀었는데, 그건 환불했어.
22. 옷을 꼼꼼히 살펴본 줄 알았는데 집에 와서 다시 보니까, 얼토당토않게 바지에 큰 구멍이 나 있더라고.
23. 너무 속상하더라. 남아 있는 같은 바지 중에는 내 사이즈가 없길래, 그냥 돈으로 돌려받았어.
24. 그래도 정말 다행이었던 건, 내가 영수증을 갖고 있었다는 거야. 그 덕분에 바로 환불할 수 있었거든.
25. 다른 디자인으로 교환할까 고민도 해 봤는데, 하필 나한테 안 어울리는 물방울무늬밖에 없더라고.
26. 요리하다가 손을 좀 베었는데, 피가 멈추지 않아서, 일단 휴지로 싸서 왔어요. 지금도 많이 아파요.
27. 피가 계속 날 정도로 많이 다쳤으면, 병원에 가시는 게 좋습니다. 심하지는 않은 것 같은데, 한번 볼까요?
28. 베인 것 같은데, 하루에 한두 번 상처에 바를 약을 주시면 좋겠어요. 소독은 안 해도 될까요?
29. 이 연고가 소독까지 되기 때문에, 이것만 바르면 괜찮으실 겁니다. 연고 값은 삼천 원(3,000원)입니다. 계산해 주세요.
30. 한 일주일(1주일) 전부터 얼굴에 자꾸 뭐가 나는데, 깨끗하게 세수를 해도 좋아지지 않네요. 약을 좀 바를까요?
31. 혹시 최근에 화장품을 바꾸거나, 스트레스를 받는 일이 있었습니까? 회사 일이 많아서, 늦게 주무시는 것은 아닌가요?
32. 제 생각에는 여드름인 것 같습니다. 중고등학교 다닐 때에는 여드름이 안 났는데, 지금 날 수도 있나요?
33. 네, 청소년들만 나는 것이 아니라, 나이가 들어서도 충분히 생길 수 있습니다. 약을 처방해 드릴 테니, 드셔 보세요.
34. 어떻게 복용하시냐면, 약은 하루에 세 번 드시고, 연고는 다른 것과 같이 사용하지 말고, 이것만 얇게 펴 바르세요.
35. 안색이 많이 안 좋아지고, 열이 나는 것 같아서, 해열제를 먹고 쉬었는데, 괜찮아지지 않는 것 같아요.
36. 어제 저녁을 먹은 후부터, 윗배가 아프고, 소화가 잘 안 되는 느낌처럼 답답해요. 계속 안 좋은데요.
37. 배탈이 났으면 설사를 했을 텐데, 화장실은 잘 가셨어요? 아래쪽 배가 특별히 아프시진 않으세요?
38. 어제 먹은 음식 때문에 체한 모양인데, 급하게 드셨어요? 빨리 드시거나, 충분히 씹지 않고 넘긴 것은 아닌가요?
39. 음식이 체했을 때에는 위가 제대로 운동을 하지 않기 때문에, 더욱 조심해야 합니다. 가능하면 오늘 음식을 삼가세요.
40. 물도 마시면 안 돼요? 따뜻한 보리차나 생수를 끓여서 마시면, 소화 불량에 도움이 되지 않을까요?
41. 자꾸 재채기가 나서 왔는데, 왜 그런지 모르겠어요. 이전에도 이런 적이 있었어요. 이번이 처음이 아니에요.
42. 매년, 사월(4월)이나 오월(5월)쯤이 되면 이런 것 같은데, 외출을 안 하고 쉬어도 문제가 해결되지 않아서, 걱정입니다.
43. 같은 현상이 주로 봄에 나타난다면, 꽃가루 알레르기일 가능성이 높습니다. 약을 드셔야겠네요.
44. 사실 알레르기는, 사람마다 고유한 체질 때문에 나타나는 것이라서, 쉽게 치료되기는 어렵습니다.
45. 약이 진정 효과를 주는 경우가 많이 있으니, 복용해 보시고, 되도록이면, 공기가 나쁜 계절엔 마스크를 착용하세요.
46. 의사 선생님, 저는 운동을 하다가 발목을 삐었는데, 걸을 때마다 아직도 아파요. 사진을 찍어 볼까요?
47. 무릎 아래쪽부터 발끝까지 부은 걸 보니까, 부러졌을 가능성이 있어 보입니다. 며칠 동안 못 걸을 수도 있어요.
48. 사진을 봐야 알겠지만, 골절된 경우에, 붕대를 감고 목발을 짚고 다니면, 다 나을 때까지 보통 한 달 걸려요.
49. 시간이 날 때마다, 아픈 부위 위에 얼음찜질을 해 주면, 통증을 가라앉히는 데에 도움이 되니, 오늘부터 해 보세요.
50. 다 나은 것 같다고 해서 절대로 다시 운동을 하시면 안 되고, 병원에 다시 방문해서, 확인을 받는 것이 좋습니다.
51. 안녕하세요. 오늘 오후 한 시(1:00)쯤 부산에 가는 표 있나요? 조금 급하게 부산에 내려갈 일이 생겨서요.
52. 네, 손님. 우선 어느 기차를 타고 싶으세요? 우리 역에는 케이티엑스(KTX)와 새마을호, 무궁화호 이렇게 기차 세 종류가 있습니다.
53. 제가 한국에서 기차 타는 게, 이번이 처음이라 잘 모르겠어요. 기차 세 가지가 어떻게 다른가요?
54. 케이티엑스(KTX)는 가장 빠르지만 비싸고, 무궁화호는 가장 싸지만 느립니다. 새마을호는 그 중간입니다.
55. 서울에서 부산까지 가는 데 얼마나 걸리나요? 세 시간 이상 걸려요? 오래 걸리지 않으면 좋겠어요.
56. 케이티엑스(KTX)는 두 시간 반 걸리고요, 무궁화호는 다섯 시간(5시간) 걸리고, 새마을호는 네 시간(4시간) 정도 걸립니다.
57. 그럼 제일 빠른 케이티엑스(KTX)를 타야겠어요. 제가 일이 좀 바빠서요, 근데 케이티엑스(KTX) 표는 얼마인가요?
58. 일반실과 특실 중, 고를 수 있는데 어디로 하시겠어요? 특실은 좌석이 넓어서 편하고, 사람도 적습니다.

59. 특실이 아무래도 좋겠네요. 몸이 피곤해서 편하게 가고 싶은데, 일반실에 비해 특실은 많이 비싼가요?
60. 아시다시피, 특실이 일반실보다 물론 더 비쌉니다. 특실은 팔만 삼천 원(83,000원)이고, 일반실은 육만 칠천 원(67,000원)입니다.
61. 역시 특실 가격이 좀 비싸네요. 아쉽지만 저는 그냥 일반실로 할게요. 일반실도 깨끗하지요?
62. 네, 손님. 죄송하지만, 한 시(1:00) 기차는 남은 좌석이 없습니다. 두 시(2:00)하고 아홉 시(9:00) 표가 있는데 언제로 하시겠습니까?
63. 그러면 두 시(2:00) 기차로 예약해 주세요. 이왕이면 창가 자리면 좋겠어요. 근데 오늘 왜 이렇게 자리가 없나요?
64. 아마 주말이라 사람들이 부산으로 많이 내려가기 때문에, 예약이 모두 찬 것 같습니다. 현금 결제신가요?
65. 그렇군요, 기차 요금은 현금으로 결제할게요. 이제 이 표를 가지고 어디로 가서 타면 되나요?
66. 우선 대합실에 앉아 계시다가, 화면에 열차 번호와 시간을 확인하시고요, 기차가 도착하면 승강장으로 내려가시면 됩니다.
67. 조금 이따가 승강장으로 어떻게 내려가나요? 엘리베이터를 타나요, 아니면 계단을 이용해야 되나요?
68. 잠시만요, 손님 타는 곳이 삼번(3번)이니까 삼번(3번) 출구 왼쪽으로 나가서, 그 계단으로 내려가시면 돼요.
69. 기차를 탈 때, 기차표를 누구한테 보여 줘야 해요? 아니면 타기 전에, 무인 기계에 찍고 타나요?
70. 아니요. 표는 나중에 승무원이 확인하니까, 손님은 케이티엑스(KTX) 승강장 이십오번(25번)을 찾아서 오호차(5호차)를 타시면 됩니다.
71. 기차 안에 매점이나 자판기가 있나요? 기차를 오래 타는데, 가는 동안에 배고플 것 같은데요.
72. 칠호차(7호차)에 간식 자판기가 있습니다. 하지만 종류가 많지는 않아서, 기차 타기 전에 미리 사서 가는 게 좋아요.
73. 실례지만, 제가 서울역을 잘 몰라서요. 혹시, 이쪽 근처에 간단하게 먹을 것 살 곳이 있나요?
74. 앞으로 직진하다 오른쪽으로 가면 편의점도 있고, 가래떡이나 식혜를 살 수 있는 가게도 있어요.
75. 저기 가게가 보여요! 기차를 놓치지 않게 서둘러서 가 봐야겠어요. 도와주셔서 정말 감사합니다.

Hye-Won Choi
Professor
Department of English Language and Literature
Ewha Womans University
52, Ewhayeodae-gil, Seodaemun-gu,
Seoul, 03760 Korea
E-mail: hwchoi@ewha.ac.kr

Gunhee Ko
Graduate Student
Department of English Language and Literature
Ewha Womans University
52, Ewhayeodae-gil, Seodaemun-gu,
Seoul, 03760 Korea
E-mail: gunheeko@ewhain.net

Received: 2023. 02. 20.
Revised: 2023. 08. 07.
Accepted: 2023. 08. 09.