Check for updates

Phonological processes of consonants from orthographic to pronounced words in the Seoul Corpus*

Byunggon Yang**

Dept. of English Education, Pusan National University

Abstract

This paper investigates the phonological processes of consonants in pronounced words in the Seoul Corpus, and compares the frequency distribution of these processes to provide a clearer understanding of conversational Korean to linguists and teachers. To this end, both orthographic and pronounced words were extracted from the transcribed label scripts of the Seoul Corpus. Next, the phonological processes of consonants in the orthographic and pronounced forms were tabulated separately after syllabifying the onsets and codas, and major consonantal processes were examined. First, the results showed that the majority of the orthographic consonants' sounds were pronounced the same way as their pronounced forms. Second, more than three quarters of the onsets were pronounced as the same forms, while approximately half of the codas were pronounced as variants. Third, the majority of different onset and coda symbols were primarily caused by deletions and insertions. Finally, the five phonological process types accounted for only 12.4% of the total possible procedures. Based on these results, this paper concludes that an analysis of phonological processes in spontaneous speech corpora can improve the practical understanding of spoken Korean. Future studies ought to compare the current phonological process data with those of other languages to establish universal patterns in phonological processes.

Keywords: phonological processes, consonants, orthographic, pronounced, Seoul Corpus

1. Introduction

A group of Korean scholars built a spontaneous Korean speech corpus that could be comparable to an American English speech corpus in the number of participants and the content of the data composition (the Buckeye Corpus by Pitt et al., 2007; the Seoul Corpus by Yun et al., 2015). The Seoul Corpus provides linguists and phoneticians much information about various aspects of the Korean language, especially the phonological processes of spoken Korean. The researchers transcribed the spoken texts using their own phonetic symbols, incorporating the unique orthographic consonant clusters in Korean. This paper adopts these phonetic symbols in its descriptions.

The Korean language consists of nineteen consonants (Sohn, 1999:153). Twelve stops are divided into lax, aspirated, and tensed categories. All of the stops are voiceless, but lax stops are lightly voiced between voiced sounds in fast and casual speech. The Seoul Corpus did not indicate the voicing distinction with different

symbols. Sohn (1999) noted that the tensed stops are equivalent in quality to the English voiceless stops that occur after /s/. He also noted that the alveodental consonants are produced with the tongue touching or approaching the back of the upper or lower teeth, which is different from alveolar English stops. In Korean, there are three fricatives: aspirated /s0, hh/, and tensed /ss/. In addition, there are three nasals⁻/mm, nn, ng/⁻and one lateral or flap sound. Sohn introduced the major phonological processes in Korean. Among them, the coda neutralization rule allows only seven consonants⁻/k0, nn, t0, ll, mm, p0, ng/⁻in the codas. The h-aspiration rule applies to produce an aspirated stop when the initial and final lax stops are merged with the previous coda /hh/ or the following onset /hh/. Similarly, the lateralization rule converts the nasal /nn/ into the lateral /ll/. In the nasalization rule, all non-nasal consonants other than the lateral become nasal before a nasal consonant. The tensification rule changes the lax stops and the fricative /s0/ to their tensed counterparts after the preceding coda stops. Finally, the palatalization rule refers to the process by which alveodental stops become palatal stops.

Jung (2019) examined consonant minimal pairs in the onset positions in a Korean dictionary and reported that /k0, s0, c0, p0, ch/ were the major phonemes forming the minimal pairs in order of greatest frequency. The velar stop /k0/ was recorded as the most frequent phoneme (12.1%), while the least frequent phoneme was /pp/ (1.1%). The aspirated affricate /ch/ appeared at 8.3%⁻almost double the proportions of the labial stop /ph/ and alveodental stop /th/. Among the obstruents, the lax consonants created more minimal pairs than the aspirated and unaspirated consonants, while the nasal consonants /mm, nn/ created more pairs than the lateral /ll/ did. These results indicate that even individual onset consonants function differently in the Korean sound system. Jung added that the unaspirated phonemes /pp, cc, ss, kk, tt/ were employed sparingly to create minimal pairs. The five phonemes totaled only 6.8%. From this finding, she proposed a teaching hierarchy of minimal pair sets of Korean consonants in the order of the most frequent /c0-ch/ pair, followed by the /p0-ph/ and /t0-th/ pairs. The least common pair was /kk-kh/, which could be supported by an analysis of authentic use in the spontaneous speech corpus. In other words, a simple frequency analysis of onset phonemes alone might not reflect the whole function of syllable components altogether. Some onsets are phonologically created by the preceding codas in Korean. The authentic use of phonemes in the spontaneous speech corpus might have to be reflected in the establishment of the teaching hierarchy, let alone the analytic results from the Korean dictionary.

Sohn et al. (2016) investigated phonetic realization of aspiration mergers in the Seoul Corpus. They found that aspiration was gradiently realized from a detailed Voice Onset Time (VOT) analysis of the stops. They also reported that age played an important role in the gradient phonetic realization of the aspiration merger. In the analysis, they extracted the two codas⁻/k0, p0/⁻before the following onset fricative /hh/, including the age and sex information of the participants. They classified the merger into six categories by both measuring the acoustic VOT durations and listening to the phrases. They examined 349 instances of the merger environments: 307 /k0/s and 42 /p0/s. Approximately half of them were judged to be aspirated, but the other half were not. When the data were broken down into the four age groups, the tens and thirties groups realized the merger more often than half of the time, but only 38% of the twenties and forties groups realized the merger. Their study is intriguing because they examined the phonological processes

in the spontaneous speech corpus, which might contribute greatly to the establishment of practical lesson plans for Korean.

Yang (2016) presented a contingency table of the orthographic and pronounced phonemes of syllable onsets and codas in the Seoul Corpus. The table delineates a general picture of the distribution of syllable components, but a one-to-one comparison of the orthographic and pronounced forms considering the adjacent syllables still requires further analysis focused on the phonological processes. One way to match the syllable components could be to equalize the number of syllables of both the orthographic and pronounced symbols. In addition, Yang (2017) examined the vowel frequency distribution of the Seoul Corpus. Of the 546,404 vowels, 90.5% were pronounced as the same orthographic symbols, while 9.5% were realized as different symbols. For example, the vowel /aa/ appeared in 114,205 instances, and 98.7% of them were pronounced as the same vowel. In addition, 34.2% of the phrases showed the same symbols, but 65.8% of them were realized as different symbols. Thus, Yang claimed that the vowels were robust in the phonological processes and noted that the majority of phonological processes of variants might be derived from the consonants, and he recommended studying the consonantal variants further.

The main purpose of this study was to contribute to the understanding of the Korean phonological processes of the Seoul Corpus by analyzing complex consonantal processes. Specifically, the current study was designed to investigate distributions of the phonological processes of Korean consonants in the Seoul Corpus by onsets and codas, as well as to explore the frequency distributions of major phonological processes.

2. Method

2.1. The Seoul Corpus

The Seoul Corpus consists of forty speakers (Yun et al., 2015 for details). Each speaker participated in an individual interview session of one hour. The spontaneous speech was performed regarding his or her own and family matters, including neighbors' concerns, leisure activities and controversial issues. The sound files were transcribed using phonetic symbols for both orthographic and pronounced texts. All of the participants were Seoul speakers. The orthographic symbols refer to an exact transcription of the written text, while the pronounced symbols refer to the phonetic transcription of the spoken text. The Korean consonants were transcribed by 30 symbols, including such consonant clusters as /ks, lm, ps/. The first symbols of the following list of triads are used in this paper: p0-ㅂ -p; s0-ㅅ-s; ph-ㅍ-pʰ; ss-ㅆ-s* ; pp-ㅃ-p*; hh-ㅎ-h; t0-ㄷ-t; c0-ㅈ -tɕ; th-ㅌ-tʰ; ch-ㅊ-tɕʰ; tt-ㄸ-t*; cc-ㅉ-tɕ*; k0-ㄱ-k; mm-ㅁ-m; kh-ㅋ -kʰ; nn-ㄴ-n; kk-ㄲ-k*; ng-ㅇ-ŋ; ll-ㄹ-l. The second and third symbols in the list indicate the Korean alphabets and the IPA symbols.

2.2. Data collection and analysis

Data for analysis were collected in two steps: extraction of participants' information and phonetic symbols; and tracing of the phonological processes.

First, the file names and orthographic and pronounced phrases of the phonetically transcribed text files were extracted from the Seoul Corpus using Praat software (Boersma & Weenink, 2020). The phrasal unit is based on the Korean orthography. The total number

of orthographic and pronounced phrasal pair rows in the integrated file was 231,632.

Second, the integrated file was imported into R Studio software, and an R script was created to trace the phonological processes from the orthographic to the pronounced phonetic symbols of English consonants (R Core Team, 2020). In the Seoul Corpus, each syllable is separated by a hyphen, and two letters are assigned to each phoneme. The syllabification was conducted by counting the number of letters of each syllable. The script checked the first two-letter phoneme as a possible onset or peak from the list of 30 consonant (cluster) symbols and 21 vowel symbols and assigned each syllable component to one after another. For example, a six-letter syllable was exactly matched with the three syllable components. The phrase /llxxmm/ was divided into the onset /ll/, the peak /xx/, and the coda /mm/. The first two-letter phoneme in a four-letter phrase /xxmm/ was assigned to the peak of the syllable and then to the coda /mm/. The same procedure was applied to the pronounced phrase on the same pair row. The number of vowels in a word was used to set a looping frequency in the script.

The orthographic syllable components were matched to the pronounced syllable components on the same pair row. Then, the current onsets and codas, as well as the onsets of the subsequent syllable, were extracted to define specific phonological environments. A template matrix with one row and fourteen columns was created to store the analysis outputs. For example, an output matrix read, "2, same, ll>ll, xx>xx, diff, mm>, mm#>#mm, xx-mm#-xx>xx-#mm-xx, s01m16f1". The partial output indicated the result of the second syllable /llxxmm/. The first column traced the syllable number, and the second column represented the onset comparison of the orthographic and pronounced forms as the same, while the fifth column indicated the coda comparison as different. The latter two columns provided information about the deletion of the coda /mm/, which moved into the onset of the following syllable. The matrix was iteratively appended to the output file, including the phonological process information. When the syllable number of the orthographic symbols of a given phase was different from that of the pronounced symbols, the author manually checked them and added dummy vowels to equalize the number of the syllables. In this way, matching of the onsets and codas of the two different sets of symbols was made possible to trace the relevant processes. Finally, the output text file of the syllable components was processed using another R script to table the frequency distributions of the same or different phonemes between the orthographic and pronounced symbols, as well as the patterns and environments of phonological variants. A third R script was created to use the patterns and environments and determine the frequency sum of such major phonological processes as aspiration, nasalization, neutralization, palatalization, and tensification and to explore how often these processes are realized in daily conversations.

3. Results and Discussion

3.1. Phonological processes of onsets

Table 1 lists the frequency distribution of the phonological processes of the onsets of the orthographic and pronounced forms in the Seoul Corpus. The total number of onsets is 467,510, consisting of at least one onset in either orthographic or pronounced forms. The types include the same symbols for the orthographic and pronounced forms and different symbols for the two forms.

Table 1. Frequency distribution of onsets in the Seoul Corpus by the agreement types (same or different symbols) of the orthographic and pronounced forms

Types	n	%
Same	362,120	77.5
Different	105,390	22.5
Total	467,510	100.0

Generally, the same symbols prevail at 77.5%, which might be related to the importance of the onsets in word identification. From a study of onsets in the Buckeye Corpus, Yang (2019) found that the majority of English onsets were pronounced as the same orthographic symbols and even with the same voicing quality. He attributed the higher rate of preservation to the speakers' strategy of delivering clearer messages to the listeners. The onsets play important roles leading to the identification of any given word compared with the codas. Cutler (1982) mentioned that the beginning of a word is the most salient part for identification. As soon as an onset is provided, the listeners can immediately narrow down the number of possible vocabulary candidates within their language system. The more onsets that are given, the smaller that the set size of the candidates can become. Yang (2018) also pointed out in his paper on the Buckeye Corpus that speakers tended to keep vowels that were important to convey their thoughts, but they tended to change or delete vowels that were relatively unimportant to deliver their intended message.

Table 2 lists the phonological processes of onsets from the orthographic forms to their pronounced forms in the Seoul Corpus. The list includes only cases with more than 50 instances to render the table simpler. The seven most common consonant insertions were added to the end of the table. The same screening criteria and additions were applied to create a coda table in the next section.

Table 2. Phonological processes of onsets in the Seoul Corpus. The column name *Orth* represents the orthographic forms, while *Pron* stands for the pronounced phonetic symbols of Korean consonants. *n* is the number of occurrences, and ∅ represents null.

Orth	Pron	n	Orth	Pron	n	Orth	Pron	n
c0	c0	39,313	kh	kh	2,098	pp	pp	1,082
c0	cc	5,215	kh	kk	61	s0	s0	32,994
c0	ch	578	kk	kk	6,795	s0	ss	4,060
c0	∅	106	kk	k0	212	s0	∅	121
cc	cc	2,130	ll	ll	32,076	ss	ss	1,673
ch	ch	7,696	ll	∅	3,746	t0	t0	40,085
hh	hh	15,039	ll	nn	285	t0	tt	5,602
hh	∅	10,406	ll	hh	141	t0	th	370
hh	nn	1,898	mm	mm	29,938	t0	∅	133
hh	kh	1,521	mm	∅	207	t0	nn	115
hh	ll	1,251	nn	nn	40,266	th	th	4,782
hh	th	594	nn	∅	2,992	tt	tt	6,768
hh	mm	500	nn	ll	453	tt	th	241
hh	ph	352	nn	mm	395	∅	nn	11,499
hh	k0	74	nn	t0	90	∅	ll	8,155
k0	k0	80,151	p0	p0	15,710	∅	ss	6,008
k0	kk	8,990	p0	pp	867	∅	k0	5,867
k0	∅	6,615	p0	∅	55	∅	mm	3,898
k0	kh	3,367	ph	ph	3,524	∅	th	2,891
k0	hh	107	ph	pp	68	∅	p0	1,465

Table 2 shows that the majority of the onset consonants were pronounced as the same orthographic symbols, as seen in Table 1. The insertions list 9.2% (42,786 instances) of the total possible onset processes (467,510); 8.2% (38,177) of the onsets are pronounced as different phonemes. The deletions list 5.2% (24,427). When we combine the instances of deletions with those of insertions, 14.4% (67,213) account for the different onsets. Thus, we can say that the majority of the pronounced onset variants can be traced to the insertions and deletions in the onsets. Specifically, according to the manner categories, the same symbols of the stops account for almost 44.5% of the onsets, followed by 19.4% nasals and 13.7% fricatives and 13.6% affricates. The lateral /ll/ amounts to 8.8%. When the onsets are divided into the lax, aspirated and tensed categories, including the fricatives /s0, ss/, the lax onsets account for 85.1%, followed by the aspirated and tensed onsets, which are almost comparable at 7.4% and 7.5%, respectively. Individually, the velar stop /k0/ is recorded the most frequent phoneme with 80,151 instances of the same processes, followed by the nasal /nn/ and alveodental stop /t0/. Yang (2016) reported the highest frequency (19.4%) of /k0/s in the 443,491 pronounced symbols in the Seoul Corpus. All of these distributions in Korean are nonlinear.

Tensed onsets are mostly preserved and pronounced with a limited number of variants. For example, there are 6,795 /kk/s, 6,768 /tt/s, 2,130 /cc/s, and 1,673 /ss/s. The majority of /tt/s were pronounced as the same symbols, while only a fraction of the symbols became the variant /th/s. This finding might be related to the distinct acoustic features of the tensed consonants. The VOT of the tensed consonants in Korean are very short compared with those of the lax and aspirated consonants (Lisker & Abramson, 1964). In addition, if the tensed onsets became variants, the word would be difficult to identify in the Korean vocabulary.

The fricative /hh/ has more than nine variants, indicating that the fricative is not very robust in the phonological processes. The deletion rate of /hh/ records the most, with 10,406 instances, accounting for 42.6% of the total onset deletions including symbols with fewer than 50 instances. The aspirated coda /hh/s are included in the count. The high onset /hh/ deletions are related to weakening of the syllable in the middle of the phrase. For example, the orthographic form /oo-hhii-llyv/ was pronounced as /oo-ii-llyv/. The listener might not have any difficulty guessing the original message without the sound /hh/ in this case, perhaps related to the low energy in the waveform. However, a majority of the other fricative /s0/s with low intensity were pronounced as the same form, perhaps related to the spectral information with the sensitive frequency band of approximately 7,000 Hz. The human ear is quite sensitive to this frequency region, which is the second resonant frequency in the 2.5-cm-long ear tube. The minimum audible field represents the region with a very low absolute threshold in the map (Moore, 2001, Figure 2.1). In the Buckeye Corpus, Yang (2019) listed 2,751 deletions of the onset /hh/; 11,357 fricatives were realized in the pronounced forms. In contrast, the deletion of the velar stop /k0/ has 6,615 instances, which are second to the fricative sounds in the frequency distribution. The lateral /ll/ records the third largest deletion rate with 3,746 instances, almost 10% of the same lateral productions. There were also approximately 10% deletions of the laterals in the English corpus (Yang, 2019). Among the nasals, /nn/ lists 2,992 instances, but /mm/ has only 207 instances. There might be certain reasons for the deletion behavior of the two different

nasals requiring further study.

The onset /nn/ insertions are the most frequent with 11,499 instances, accounting for 26.9% of the total onset insertions. These insertions are mostly made by moving the previous codas to the current onsets. For example, the orthographic phrase /t0oonn-xxll/ was pronounced as /t0oo-nnxxll/. In addition, there are 102 rare cases of /nn/ insertions in the first syllable, as seen between the orthographic phrase /yvll-vv/ and the pronounced phrase /nnyv-llvv/. Regarding the nasal, 8,155 /ll/s and 6,008 /ss/s are inserted. The bilabial nasal /mm/ records 3,898 instances, one third of the nasal onset /nn/s. This finding shows an opposite pattern of the nasal deletions mentioned above. Thus, we can say that the majority of these different onsets in Table 1 are derived from either deletions or insertions in the Seoul Corpus.

3.2. Phonological processes of codas

Table 3 presents the phonological processes of codas in the Seoul Corpus. The total number of codas is 218,100, counting at least one coda in either the orthographic or the pronounced form. The number is smaller than half of the onsets (46.7%). From this frequency comparison, we can guess that the onsets seem more important to creating a distinctive vocabulary than the codas in the daily conversations. In addition, the frequency distributions of the two forms are almost equal, but a slightly higher rate occurred in the same symbols. The same orthographic and pronounced codas amount to 51.8%, while the variant codas amount to 48.2%. We have already seen in the previous section that the same onset symbols prevail. Thus, we could claim that the onsets and codas behave differently in the spontaneous speech corpus. The frequency distributions of consonants might have to reflect the syllable components.

Table 3. Frequency distribution of codas in the Seoul Corpus by the agreement types (same or different symbols) of the orthographic and pronounced forms

Types	n	%
Same	112,901	51.8
Different	105,199	48.2
Total	218,100	100.0

Table 4 displays the distribution of the phonological processes of codas in the Seoul Corpus. The coda deletions list 33.6% (73,323 instances), including symbols with fewer than 50 instances, which were not listed in the table, while the coda insertions list 0.9% (2,031 instances). In addition, 13.7% (29,845) of the orthographic phonemes are realized as variants. When we combine the deletion instances with those of insertion, 34.6% (75,364) of the total possible processes in Table 4 account for different codas. Here, again, the majority of the pronounced coda variants are caused by insertions and deletions. It is interesting to see more deletions in the codas compared with those in the onsets. This discrepancy could be related to the important information weight of the onset again, but the coda could be sufficiently flexible to become a variant that expedites speech production without the risk of misunderstandings by listeners in daily conversations.

According to the manner categories, the same symbols of the nasals account for almost 70.6% of the onsets, followed by 21.2% laterals and 8.2% stops. This result is different from the onset

analysis in the previous section. Here, the nasals prevail in the codas. The most preferred nasal is the alveodental /nn/, followed by the velar /ng/ and the labial nasal /mm/. Because of the coda constraints of the Korean sound system, only lax stops are in pronounced forms (Lee, 2015). Individually the alveodental nasal /nn/ is recorded as the most frequent phoneme with 45,310 instances of the same processes, followed by the lateral /ll/. There are many pronounced variants in the table. The greatest number of variants is created by the nasal /nn/: 14,064 /nn/s are pronounced as the velar /ng/; and 4,611 /nn/s are changed into the bilabial /mm/s. The tensed fricative /ss/ follows the nasal in the number of variants. The majority of the fricatives are realized as nasal /nn/s. If we sort out the analytic results by the pronounced symbols, the majority of the orthographic symbols are processed to become 26,366 nasals, which cover 88.3% of the variants. We go into detail about the major phonological process types in the next section.

Table 4. Phonological processes of codas in the Seoul Corpus. The column name *Orth* represents the orthographic forms, while *Pron* stands for the pronounced phonetic symbols of Korean consonants. *n* is the number of occurrences, and ∅ represents null.

Orth	Pron	n	Orth	Pron	n	Orth	Pron	n
c0	∅	776	ng	ng	22,782	s0	∅	2,729
	nn	106		∅	1,027		t0	238
ch	∅	227		nn	51		nn	229
	mm	57		∅	3,833		mm	136
hh	∅	5,346	nh	nn	439	ss	∅	11,651
k0	∅	11,014		ng	357		nn	2,825
	k0	7,114	nn	nn	45,310		t0	180
	ng	1,181		ng	14,064		∅	623
kk	∅	313		∅	12,699	t0	t0	156
lh	∅	207		mm	4,611		nn	130
lk	ll	92		ll	568		∅	3,958
ll	ll	23,952	p0	∅	2,076	th	nn	150
	∅	11,540		p0	1,966		t0	65
lm	ll	89		mm	765		ng	531
lp	ll	125		k0	60		ll	503
mm	mm	11,596	ph	∅	686	∅	nn	377
	∅	4,538		p0	249		t0	303
	ng	523	ps	p0	1,362		mm	170
	nn	56		mm	382		k0	75
				∅	68		p0	72

The nasal /nn/ topped the list with 12,699 instances, followed by the other codas /ss, ll, k0/ with approximately 11,000 instances. These four codas account for almost 64% of the total number of deletions. All of the other deleted codas number fewer than 5,000 except for 5,346 /hh/s. The deletion rate of the orthographic /hh/ amounts to almost half of the onset symbols in the previous section. Many of the fricatives merged with the following onset stops to become aspirated counterparts, which we describe in the following section. If we include the single fricative sound within such orthographic consonant clusters as 4,040 /nh, lh/s, the number of the specific fricative deletions might be comparable to that of the onsets. The number of insertions of the seven codas totaled 2,031. The velar nasal /ng/s topped the insertion list. Here, again, the nasals /ng, nn, mm/ in that order totaled 1,078 instances, covering 53% of the total insertion processes. As in the variant discussion above, we can say that the nasal codas are most preferred among the Korean pronounced variants and insertions. Further analysis of more

spontaneous speech corpora might be necessary to make a conclusive statement on this phenomenon.

Yang (2012, 2019) touched on the controversial issue of language universals and proposed considering the syllable components, rather than the total inventory distribution of consonants. Maddieson (1984) reported that 80% or more languages had the consonants /p, t, k, m, n, s, j/. In addition, Eckman (2004) reviewed the Markedness Hypothesis and posited that the notion of typological markedness might be closely related to language universals. From a series of studies of the English and Korean corpora, Yang (2019) urged an establishment of language universals that considers characteristics of the given category by the syllable components. We add here that these universals should reflect not only orthographic symbols found in a dictionary or written texts but also authentically pronounced symbols, which might better represent language universals.

3.3. Frequency distribution of phonological process types

This section describes the frequency distributions of five major phonological process types in the Seoul Corpus. We examine aspiration, nasalization, neutralization, palatalization, and tensification in that order. Table 5 lists the process types and frequencies and their proportions in the Seoul Corpus. Altogether, the five processes account for 12.4% of the total possible onset processes, and 467,510 can be seen in Table 1.

Table 5. The phonological process types and their frequency distributions in the Seoul Corpus. The percentage indicates the proportion of the total possible processes (467,510) in the Corpus.

Types	n	%
Aspiration	6,185	1.3
Nasalization	25,773	5.5
Neutralization	901	0.2
Palatalization	593	0.1
Tensification	24,754	5.3
Total	58,206	12.4

The total number of aspiration processes is 6,185 instances. The aspiration process accounts for 1.3% of the total possible onset processes. Four variants of aspiration in the pronounced outputs are recorded as 4,862 /kh/s, 601 /ch/s, 369 /th/s, and 351 /ph/s. The most frequently aspirated onset symbol occurs from 2,819 /k0/s to /kh/s after the coda /hh/ of the previous syllable. The second most frequent aspiration merger lists 1,375 instances of the same output symbol /kh/ from the coda /k0/ of the previous syllable coupled with the onset /hh/. Thus, we can say that aspiration is mostly triggered by the coda /hh/ of the previous syllable. In the introduction, we reviewed (Sohn et al., 2016) on the phonetic realization of the aspiration merger in the Seoul Corpus. They pointed out the limitation of the speech corpus to statistical representation because of insufficient utterance samples. Further resources should be collected to offer conclusive remarks on this process.

Nasalization is a type of universal phonological process. We found 25,773 instances of nasalization in the Seoul Corpus. These instances amount to 5.5% of the possible processes. Breaking them down into the nasal place categories, the most frequently nasalized coda symbol is 15,141 /ng/s, followed by 7,130 /mm/s and 3,502 /nn/s. The same pronounced onset and coda nasals appear in 7,628 instances, while the different outputs occur with 18,145 instances.

The most frequent 12,814 nasal coda output /ng/s are derived from the coda nasal /nn/s in the environment of the following onset /k0/s. The triggering context is the velar stop /k0/. If we focus on velar nasal assimilation, the process is also a type of velarization to facilitate easy and quick production of a phrase. The second and third most frequent instances are from 2,791 /ss/s to /nn/s before the onset /nn/s and from 2,014 /nn/s to /mm/s before the onset /p0/s. When the nasal /nn/ becomes the labial /mm/, the process is also a type of labialization. Among the five phonological processes in this paper, nasalization is the most frequent process, perhaps related to the fixed nasal tract, which cannot be controlled easily by such articulators as the tongue or jaw. Thus, adjacent nasals trigger the process more often than the other consonants.

There are 901 instances of neutralizations, which account for 0.19% of the possible processes − only slightly greater than the aspiration processes. Here, the same phonetic symbols from the orthographic to the pronounced forms are not included in the count because we do not want to have the small number of the variant forms hidden behind the total sum. The neutralization rule might also derive these same symbols. The distribution of the same symbols for both the orthographic and pronounced forms can be seen in the previous onset and coda tables. The majority of the neutralization is realized as the alveodental lax stop /t0/. Half of the 418 orthographic coda /s0, ss/s are pronounced as the lax stop. The other codas−/c0, ch, th/−occur at fewer than 65 instances. In addition, 29 neutralized codas are realized as the following onsets.

Palatalization is also a well-known phonological process. In the Seoul Corpus, we found only 593 instances of palatalization. The majority of them list 544 onset /ch/ outputs from the preceding coda /th/s. Only 47 preceding coda /c0/s are realized as the following onset /c0/s, constituting a very small proportion (<0.12%) of the total possible processes. Yang (2019) examined the Buckeye Corpus and reported 284 instances of English palatalization. Specifically, the palatalized /ch/ occurred in only 159 instances and 23 /jh/s and 102 /jhr/s from the orthographic cluster /dr/s in English. These palatalization processes are found to be rare phonological instances in authentic English and Korean corpora. Many linguists would introduce the process quite often in a regular linguistics course, but we should consider whether this phenomenon should be placed higher in the teaching hierarchy of Korean lesson plans.

In the Seoul Corpus, there are five tensed output symbols. The number of tensification processes is 24,754, accounting for 5.3% of the total possible processes. The most frequent case is from the 2,549 coda /k0/s to the tensed onset /kk/s in the environment of the onset /k0/. In addition, 2,377 coda /ss/s were combined with the following onset /t0/s to become the tensed onset /tt/s, while 2,331 coda /ss/s were combined with the following onset /k0/s to be pronounced as the tensed onset /kk/s. We observed that the orthographically tensed onsets are pronounced as the same sounds in the previous section. The tensed consonants in the order of the most favored pronounced symbols are 9,000 /kk/s, 5,602 /tt/s, 5,216 /cc/s, 4,069 /ss/s, and 867 /pp/s. Velar tensification is most prevalent in the coda.

So far, we have examined five individual phonological process types and found that a relatively small proportion of the processes are involved in the pronounced forms. This finding might be related to how the participants spoke in the interviews. They might have produced their speech rather in a clearer speech mode. Listening to the output also confirms this impressionistic notion. We might claim in haste that Korean people tend not to use these phonological processes very much in daily conversations. Rather, we would like to point out that the Korean participants applied a relatively small number of phonological processes in the conversations to sound clearer. We would expect many more phonological processes if we had a spontaneous corpus of telephone conversations or casual speech modes. In English, Patterson & Connine (2001) found more flappings of a medial /t/ sound in intimate telephone conversation. Additionally, Ernestus et al. (2015) reported that a higher percentage of semantically weak variants emerged in casual face-to-face and telephone conversations.

4. Summary and Conclusion

This study examined the phonological processes of Korean consonants in the Seoul Corpus. R scripts were created to syllabify each word in both orthographic and pronounced forms. Then, the one-to-one phonological processes of English consonants were tabulated after equalizing the number of syllables in both forms, and they were divided into such syllable components as onsets and codas separately. In addition, the frequency distributions of five phonological processes were examined in detail.

The results are as follows. First, the majority of the consonants were pronounced as the same sounds in conversation. The participants in the Seoul Corpus must have attempted to be more clearly understood, as reported in the American Buckeye Corpus. Second, more than three quarters of the onsets are pronounced as the same forms, while approximately half of the codas are pronounced as variants. This finding might be related to the relatively important role of the onsets in the identification of given words. We observed a similar trend in the preserved tensed onsets. Third, the major instances of different onset and coda symbols are mostly by deletions and insertions. The insertions of consonants are mostly performed in the onsets and are very limited in the codas. The majority of the coda insertions are nasal variants before the subsequent nasal onsets. Finally, the five phonological process types accounted for only 12.4% of the total possible procedures, excluding the same symbol outputs.

From these results, the author concludes that an analysis of phonological processes in spontaneous speech corpora can improve the practical understanding of spoken Korean. Further studies would be desirable to compare the current phonological process types with those of other languages to identify any universal patterns.

References

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.fon.hum.uva.nl/praat/

Cutler, A. (1982). The reliability of speech error data. In A. Cutler (Ed.), *Slips of the tongue and language production* (pp. 7-28). Amsterdam, Netherlands: Mouton.

Eckman, F. R. (2004). Universals, innateness and explanation in second language acquisition. *Studies in Language*, *28*(3), 682- 703.

Ernestus, M., Hanique, I., & Verboom, E. (2015). The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics*, *48*, 60-75.

Jung, J. (2019). A study of phoneme frequencies based on the

pronunciation in the entries of Korean dictionary. *Language and Linguistics*, *83*, 179-218.

Lee, M. G. (2015). Modern Korean phonology for Korean education. Seoul: Hankook.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements, *Word*, *20*(3), 384-422.

Maddieson, I. (1984). *Patterns of sounds.* Cambridge: Cambridge University Press.

Moore, B.C. (2001). *An introduction to the psychology of hearing.* San Diego, CA: Academic Press.

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production. *Phonetica*, *58*, 254-275.

Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. (2007). *Buckeye corpus of conversational speech* (2nd ed.). Columbus, OH: Department of Psychology, Ohio State University. Retrieved from https:// buckeyecorpus.osu.edu/

R Core Team (2020). R: A language and environment for statistical computing (version 3.6.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https:// www.R-project.org/

Sohn, H., Lim, S., & Ahn, M. (2016). Phonetic realization of aspiration merger of Korean stops: A spontaneous speech corpus-based study. *The Journal of Linguistic Science*, 78, 189-213.

Sohn, H. M. (1999). *The Korean language.* Cambridge, UK: Cambridge University Press.

Yang, B. (2012). Reduction and frequency analyses of vowels and consonants in the Buckeye Corpus. *Phonetics and Speech Sciences*, *4*(3), 75-83.

Yang, B. (2016). Phoneme distribution and phonological processes of orthographic and pronounced phrasal words in light of syllable structure in the Seoul Corpus. *Phonetics and Speech Sciences*, *8*(3), 1-9.

Yang, B. (2017). Phonological processes of vowels in pronounced phrasal words of the Seoul Corpus by gender and age group. *Phonetics and Speech Sciences*, *9*(2), 23-29.

Yang, B. (2018). Phonological processes of vowels from orthographic to pronounced words in the Buckeye Corpus by sex and age groups. *Phonetics and Speech Sciences*, *10*(2), 25-31.

Yang, B. (2019). Phonological processes of consonants from orthographic to pronounced words in the Buckeye Corpus. *Phonetics and Speech Sciences*, *11*(4), 55-62.

Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., ... Kim, J. (2015). The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences*, *7*(2), 103-109.

• **Byunggon Yang,** Corresponding author
Professor, Dept. of English Education
Pusan National University
30 Changjundong, Keumjunggu
Busan 46261, Korea
Tel: +82-51-510-2619
Email: bgyang@pusan.ac.kr
Homepage: http://fonetiks.info/bgyang
Fields of interest: Phonetics, Phonology