Linguistic Research 30(2), 313-333

Schwa deletion in the conversational speech of English: The role of linguistic factors*

Na-Young Ryu · Sung-Hoon Hong**
(Hankuk University of Foreign Studies)

Ryu, Na-Young and Sung-Hoon Hong. 2013. Schwa deletion in the conversational speech of English: The role of linguistic factors. *Linguistic Research* 30(2), 313-333. In this study we investigate schwa deletion and the role of three linguistic factors based on the Buckeye Speech Corpus (Pitt *et al.* 2007). This study provides evidence that schwa deletion is strongly influenced by stress environment *(pre-stress position vs. post-stress position)*, sonority distance between two consonants surrounding an unstressed vowel [ə], and lexical frequency of the word containing the schwa. Stress environment affects schwa deletion in such a way that schwa is more likely to delete in post-stress position than in pre-stress position. As for the effect of sonority, schwa deletion is more likely to occur when there is a greater difference in sonority between the two consonants that appear before and after the schwa. With respect to word frequency, schwa deletion is less frequent in low frequency words than in high frequency words. We conduct a logistic regression analysis to find out how influential these factors are. The results of the regression analysis show that stress environment is the most pervasive, sonority the next, and lexical frequency the least influential factor that affects schwa deletion. **(Hankuk University of Foreign Studies)**

Keywords schwa deletion, stress, sonority, lexical frequency, the Buckeye Speech Corpus

1. Introduction

Speech variation has received a great deal of attention since studies on conversational speech became more common. It has been shown that the amount of reduction and deletion in spontaneous speech is greater than expected (Johnson 2004). Variable schwa deletion in the conversational speech of English is currently a challenging topic, since it manifests complex interactions and relationships between

* We appreciate anonymous reviewers for their helpful feedback. This work was supported by Hankuk University of Foreign Studies Research Fund of 2012.
** Corresponding author.

linguistic and extra-linguistic factors.

Previous research has provided some insights into the phenomenon of schwa deletion. Many studies have confirmed that schwa deletion is influenced by multiple factors such as lexical stress position, sonority, lexical frequency, word length, phonotactic environment, and speech style.

Among these, three factors stand out. Firstly, Hooper (1976, 1978) points out that schwa deletion exhibits different behaviors depending on whether the schwa is placed before or after a stressed syllable. Patterson *et al*. (2003), based on their study of the Switchboard speech corpus (Godfrey *et al*. 1992), also reveal that the most significant factor affecting schwa deletion is the stress environment. Secondly, a number of researchers observe that vowel deletion readily occurs when the sonority difference between the consonants neighboring schwa is great enough. In other words, the greater the sonority difference between the consonants surrounding the target vowel, the more the vowel is likely to delete. Zwicky (1972) also notes that the sonority requirement, although not categorical, affects schwa deletion positively.

In addition to the effects of stress environment and sonority, many researchers suggest the role of lexical frequency in vowel reduction or deletion. Fidelholtz (1975), for example, points out that unstressed vowels in higher-frequency words are more likely to reduce than those in the lower-frequency words (e.g. *[ə]strónmy* 'astronomy' vs. *g[æ]strónomy* 'gastronomy'). Similarly, Fokes and Bond (1993) and Bybee (2000) also suggest that frequent words readily undergo vowel deletion in English.

As mentioned above, although a large number of researchers have claimed that the three linguistic factors play a role in schwa deletion, empirical evidence is limited in the sense that the previous studies were conducted based on small-scale databases. In this study, we utilize the Buckeye Speech Corpus (BSC; Pitt *et al.* 2007)[1], a large-size corpus containing more than 300,000 words of spontaneous American English speech, to examine how these factors affect the distribution of schwa absence and presence. As a follow-up study of Ryu and Hong (2011, 2012), this study will contribute to elucidating the patterns of schwa deletion by conducting statistical analyses about how linguistic factors influence the variable deletion process

[1] In the BSC, the recorded speech is orthographically transcribed and phonetically labeled. The phones are automatically transcribed and then corrected manually by phonetically trained research assistants (Kiesling *et al*. 2006).

and which of these factors is most significant in determining schwa-deleted word forms.

The organization of this paper is as follows. Section 2 provides a description of the data, along with a general overview of schwa deletion. Section 3 presents the background on schwa deletion, paying a particular attention to linguistic factors. Section 4 shows the results of quantitative analyses of schwa deletion based on the BSC. The conclusion with a summary of the study follows in section 5.

2. Previous studies on schwa deletion

Schwa deletion has attracted particular interests from many researchers because both linguistic and extra-linguistic factors play a fundamental part in its occurrence.

As suggested by Zwicky (1972) and Hooper (1978), lexical stress environment has been reported as a significant factor contributing to schwa deletion. Dalby (1986) agrees that the removal of schwa is less likely in pre-stressed environment. He emphasizes that word-initial schwas in pre-stress environment (mostly two-syllable words) show a deletion rate of 44%, whereas word-medial, post-stress schwas (mostly three-syllable words) show a deletion rate of 62%. Patterson *et al*. (2003) also claim that the difference between post-stressed and pre-stressed position is predominantly responsible for an increase in the predictability of schwa deletion in conversational English. However, Glowacka (2001) points out that the position of unstressed vowels with respect to the primary word stress does not have any influence on the deletion process in read speech since the process is equally likely to occur in pre- and post-stress positions.

Besides stress, word frequency has been proposed as another factor affecting vowel reduction and deletion. Fidelholtz (1975) stresses that higher-frequency words are more likely to reduce than lower-frequency words. Hooper (1976) argues that frequent words are more likely to undergo schwa deletion than infrequent words. For example, *mem[o]ry*, which is a frequent word, tends to undergo schwa deletion, whereas *mamm[a]ry*, which is an infrequent word, does not. Fokes and Bond (1993) confirm that the application of vowel deletion is sensitive to word frequency. They observe that their highest frequency word *support* shows the highest rate of vowel deletion. Contrary to the previous research, Song (2013) shows that schwa is more likely to delete in lower-frequency words than in higher-frequency words of the Buckeye Corpus. She notes that there is no strong positive relationship between word frequency and the rate of schwa deletion.

In addition to stress and frequency, Zwicky (1972) states that the sonority requirement positively affects the vowel deletion. Hong (2007) confirms the earlier observation made by Zwicky (1972), Hooper (1976, 1978) and Pérez (1992) that the occurrences of schwa deletion are dependent on the sonority difference between the two consonants surrounding schwa.

Apart from the three factors mentioned above, several researchers suggest that phonotactics also play a role in schwa deletion. Hammond (1999), for example, shows that deletion occurs in many types of /#C_1əC_2/ environment if the deletion would result in phonotactically legal onset clusters. Patterson *et al.* (2003) confirm that all of the candidates for pre-stress schwa deletion would lead to legal onset clusters if schwa is deleted. Davidson (2006), however, insists that phonotactic environment does not have a strong effect on whether or not deletion can occur. /C_1əC_2/ sequences that would result in legal initial clusters and those that would yield illegal initial clusters demonstrate a similar extent of deletion. Glowacka (2001) also shows that when unstressed vowels in English undergo the deletion process, it leads to new types of consonant clusters, which do not conform to the intrasyllabic sonority profile. Finally, Dalby (1986) and Patterson *et al.* (2003) note that word length has an impact on schwa deletion. In their analyses, they stress that the overall deletion rate for two-syllable words is lower than three-syllable words.

Besides these linguistic factors, some extra-linguistic variables have been reported to influence schwa deletion such as speech rate, speaker’s gender, and dialect. Zwicky (1972) suggests that speakers with certain regional dialects have a greater propensity to eliminate schwa. Also, Dalby (1986) shows that schwa is deleted in 2% of the tokens in slow read speech and 44% in fast read speech. Interestingly, Patterson *et al.* (2003) manifest that neither the speaker’s sex nor the speaker’s accent has any reliable effect on the rate of schwa deletion. Lastly, Davidson (2006), based on the examination of 28 different /#CəC-/ sequences, argues that the acoustic patterns for schwa elision are consistent with a gestural overlap rather than phonological deletion.

3. Data and method

Certain vowel sounds in English are strongly associated with the absence of stress. They appear exclusively in unstressed syllables. These are known as unstressed or reduced vowels, and tend to be characterized by such features as laxness, shortness in duration, and articulation in the central vowel space (Ladefoged 2006). The occurrences of unstressed vowels [ə, ɪ, ʊ], which we obtained from the BSC, are given in Table 1.

Table 1. English unstressed vowels in the BSC

IPA	ə	ɪ	ʊ
BSC	ah	ih	uh
Number of vowels	66,425 (51%)	56,908 (44%)	5,617 (5%)

As shown in Table 1, the most common reduced vowel in this corpus is [ə]. Of the unstressed vowels, schwa [ə] accounts for 51%, and it is the prime candidate for elision in English.

For an investigation of schwa deletion, we collected 2,277 tokens with 104 word types from the BSC.[2] This data set consists of two- and three-syllable words with schwa [ə][3] (see the Appendix for the full data set). In an attempt to gather all the relevant data that conforms to potential schwa deletion environment, we conducted the following steps[4]: First, word frequency and stress have been suggested to be important in accounting for schwa deletion, but the BSC does not serve our purpose properly because it does not provide stress information and its total number of tokens is not large enough to make a stable generalization. Thus, we used the Corpus of Contemporary American English (COCA; Davies 2008)[5] to obtain word frequency and the CELEX lexical database (Baayen *et al.* 1995)[6] to acquire stress

2 In the BSC, a vowel is considered deleted when there is complete absence of voicing, formant structure, and aspiration.

3 The number of the occurrence of two- and three-syllable words in our analysis is 797 (35%) and 1,480 (65%), respectively.

4 The whole procedure was performed with Perl scripts written by the first author.

5 The Corpus of Contemporary American English (COCA) consists of more than 460 million words. The COCA offers detailed frequency data based on a wide range of genres such as spoken, fiction, popular magazines and academic texts.

information. Second, tokens containing [$C_1əC_2$] were extracted from the BSC. In order to maximize accuracy, all tokens categorized as "schwa deletion" were manually double-checked. Third, the collected tokens were divided into two groups—high- and low-frequency—based on their word frequencies.[7] Finally, among the tokens containing [$C_1əC_2$], we selected two- and three-syllable words to be analyzed in the study. After performing these steps, we obtained 2,277 tokens with schwa that were distributed across 104 different types of words.

Table 2. Number of schwa retention and deletion in the BSC

	Number of Schwa token	%
Schwa retention	1,373	60%
Schwa deletion	**904**	**40%**
Total	2,277	100%

Table 2 shows that of the 2,277 tokens containing at least one schwa, schwa retention occurs in 1,343 tokens (60%) and schwa deletion in 904 tokens (40%). This schwa deletion rate is similar to what Dalby (1986) reports for fast spoken American English. Dalby examines schwa deletion in read speech and speech from television news broadcast. In read speech, it is found that schwa is deleted in 44% of the tokens in fast speech and 2% in slow speech.

In both two- and three-syllable words containing schwa, we examined schwa deletion according to stress environment, word frequency and the sonority of the consonants surrounding schwa. Table 3 presents the occurrences of schwa deletion in some representative examples from the data set.

6 The CELEX lexical database of English is composed of 18 million tokens with a 45,838 lemma lexicon. The CELEX provides information on orthography, phonology, morphology, syntax and frequency. In particular, information on phonology includes the number of syllables and primary and secondary stress markers.

7 This binary division of frequency will be discussed in more detail in Section 4.

Table 3. Schwa deletion in some representative examples[8]

Stress environment	High frequency		Low frequency	
	Two-syl. words	Three-syl. words	Two-syl. words	Three-syl. words
Pre-stress position	*p[o]iíce* (73,350)	*t[o]mórrow* (99,675)	*c[o]mmít* (7,814)	*c[o]nnécted* (13,558)
Post-stress position	*bótt[o]m* (39,012)	*fám[i]ly* (208,789)	*vír[u]s* (11,596)	*désp[e]rate* (10,702)

We then carried out statistical analyses using IBM SPSS 20.0. The results of these analyses are given in the next section.

4. Results

In this section, we present the results of statistical analyses of schwa deletion according to the three linguistic factors and their factor values, namely lexical stress environment (pre- vs. post-stress), sonority (sonority distance between the consonants surrounding the schwa), lexical frequency (high- vs. low frequency). The distribution of the data in the BSC, arranged according to the three linguistic factors and their factor values, is presented in Table 4.

Table 4. Distribution of schwa deletion according to three linguistic factors and their factor values

Factors	Factor values	Numbers	Examples
Stress environment	Pre-stress	307	*p[o]lice*
	Post-stress	1,970	*desp[e]rate*
Sonority distance	-2	90	*terr[i]ble*
	-1	7	*par[a]noid*
	0	308	*gall[e]ry*
	1	1,053	*fin[a]lly*
	2	819	*hist[o]ry*
Lexical frequency	Low	455	*av[e]nue*
	High	1,822	*diff[e]rent*

---

[8] The values in parentheses are the token frequencies extracted from the COCA.

4.1 Effect of lexical stress environment

There are many researchers who have emphasized that lexical stress environment has an influence on schwa deletion. Zwicky (1972), for example, points out that vowel deletion occurs in two environments. One is when schwa is placed in word-initial syllable that has at least one onset consonant and followed by a stressed syllable (e.g. *s[u]ppóse*→[spoʊz]). The second environment is where schwa is located in the second syllable preceded by a syllable containing a stressed vowel (e.g. *córp[o]rate*→[kɔrprət]). Hooper (1978), Dalby (1986), Kager (1997), and Patterson *et al.* (2003) also claim that stress environment is one of the most significant factors contributing to schwa deletion in English. Patterson *et al.* (2003), in particular, explain in their corpus analysis that the most influential factor affecting schwa deletion is the stress environment. They contest that two- and three-syllable words are more likely to undergo schwa deletion in post-stressed environment than in pre-stressed environment. In Table 5, we present the relationship between the occurrence of schwa deletion and the stress environment in the BSC.

Table 5. Schwa retention vs. deletion according to stress environment

	Stress environment		Total
	Pre-stressed position	Post-stressed position	
Schwa retention	234(17%)	1,139(83%)	1,373
Schwa deletion	**73(8.1%)**	**831(91.9%)**	**904**

Table 5 indicates that the position of the unstressed vowel [ə] with respect to the main word stress has much impact on the deletion process (χ^2=37.581, p < .000). Schwa deletion is more likely to occur in post-stressed position than in pre-stressed position. In the post-stress environment, schwa tends to delete in the second syllable if it is preceded by a syllable containing a stressed vowel. In the BSC, 91.9% of schwa deletion occurs in the post-stress environment (e.g. *fam[i]ly, hist[o]ry, mem[o]ry*), whereas only 8.1% takes place in pre-stress position (e.g. *c[o]rrect, s[u]pply, t[o]day*). This pattern of schwa deletion as a function of stress environment is consistent with Patterson *et al.* (2003) and Dalby (1986), who show that the propensity of schwa deletion in pre-stressed position is low (around 9%). Also, these results largely support the findings presented in other previous studies of schwa deletion (Zwicky 1972, Hooper 1978, Kager 1996).[9]

4.2 Effect of sonority

A number of researchers have observed that the occurrence of schwa deletion is closely related to the sonority distance between the neighboring consonants (Zwicky 1972, Hooper 1976, 1978, Pérez 1992, Hammond 1999, Bybee 2000, 2001, Hong 1999, 2007). The observation is that schwa deletion in English is more likely to occur when the consonant preceding the deleted vowel is lower in sonority than the consonant following it (*cam[e]ra, p[o]lice,* and *op[e]ra*).

In order to calculate the sonority distance between the two consonants surrounding schwa, we adopt a sonority scale where numerical sonority values apply to consonants as follows: 1 to obstruents (stops and fricatives), 2 to nasals, and 3 to approximants (glides and liquids).[10]

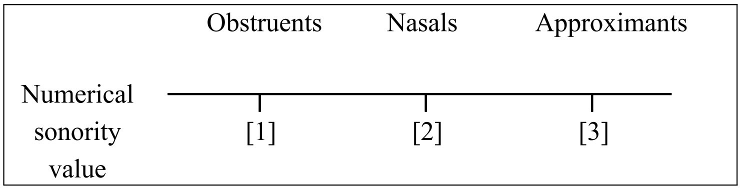

Figure 1. Sonority scale

The sonority distance is then calculated by the sonority difference between pre-schwa and post-schwa consonants, that is, the sonority value of the post-schwa consonant minus that of the pre-schwa consonant. In *cam[e]ra*, for example, the sonority distance between pre-schwa and post-schwa consonants is 1 (son(liquid *r*) − son(nasal *m*) = 3-2 = 1). The rates of schwa retention and deletion according to the sonority distance are given in Table 6.

---

9 The effect of stress environment on schwa deletion differs depending on word size (i.e. the number of syllables). In the BSC, most of pre-stress schwa deletion occurs in two-syllable words (89%; 65 out of 73 words), while post-stress deletion predominantly takes place in three-syllable words (96.4%; 801 out of 831 words).

10 More detailed sonority scales can be found in the literature. Foley (1972), for example, proposes the following scale: Vowels [6] > Glides [5] > Liquids[4] > Nasals [3] > Fricatives [2] > Stops [1]

Table 6. Effect of sonority distance on schwa retention and deletion

Sonority distance	-2	-1	0	1	2	Total
	son(C_1)>son(C_2)	←	→		son(C_1)<son(C_2)	
Schwa retention	62 (4.5%)	5 (0.4%)	232 (16.9%)	752 (54.7%)	322 (23.5%)	1,373
Schwa deletion	**28 (3.1%)**	**2 (0.2%)**	**76 (8.4%)**	**303 (33.4%)**	**495 (54.9%)**	**904**

Table 6 shows that the greater the sonority distance, the more schwa deletion is likely to occur. In other words, schwa is more likely to delete when the following consonant is higher in sonority than the preceding consonant. Chi-square analysis verifies that the sonority difference is largely responsible for the increase in the predictability of schwa deletion (χ^2 = 219.713, p < .000). This finding is in agreement with the results of the previous studies (Zwicky 1972, Hooper 1976, 1978, Pérez 1992, Hammond 1999, Bybee 2000, 2001, Hong 1999, 2007) and confirms the relationship between increase in sonority difference and schwa deletion.[11]

We further examine the deletion contexts in details. In particular, we distinguish eighteen types of [C_1əC_2] sequences and see how the examples of schwa deletion in the BSC are distributed. The frequency of each sequence type and the rate of schwa deletion are presented in Table 7.

11 Schwa deletion does not occur in equal rate in pre-and post-stressed environment. According to our study of the BSC, pre-stress deletion is favored in the context involving two obstruents (e.g, *s[u]pply, s[u]pport, t[o]day*), but no deletion is observed in the context involving two sonorants. In the post-stressed position, on the other hand, deletion is most favored in the context between O(bstruent) and S(onorant) (e.g. *bott[o]m, diff[e]rent, cath[o]lic*). Post-stress deletion takes place 68.6% in the O_S context, and 25.3% in the S_S context (e.g. *gall[e]ry, mem[o]ry, natur[a]lly*). The following table shows how schwa deletion occurs differently in two stress environments.

Context	Sonority difference	Pre-stressed position	Post-stressed position	Total
O_O	0	**38(52.1%)**	32(3.9%)	70(7.7%)
O_S	+1 ~ +2	26(35.6%)	**570(68.6%)**	**596(65.9%)**
S_O	-1 ~ -2	9(12.3%)	19(2.3%)	28(3.1%)
S_S	0	0(0%)	210(25.3%)	210(23.2%)
Total		73	831	904

Table 7. Schwa deletion according to consonant types

Sonority distance	Type of [$C_1əC_2$]	Type Freq(%)	Token Freq(%)	Examples
+2	fric_*r*	16(39%)	306(61.8%)	*av[e]rage, diff[e]rent*
	fric_*l*	9(22%)	109(22%)	*abs[o]lute*
	stop_*r*	14(34%)	68(13.7%)	*c[a]reer, op[e]ra*
	stop_*l*	2(5%)	12(2.4%)	*p[o]lice, p[o]liceman*
+1	fric_nasal	13(38%)	70(23.1%)	*av[e]nue, pers[o]nal*
	nasal_ *r*	5(15%)	10(3.3%)	*cam[e]ra, mem[o]ry*
	nasal_ *l*	3(9%)	192(63.4%)	*fam[i]ly, fin[a]lly*
	stop_nasal	13(38%)	31(10.2%)	*t[o]morrow, c[o]mmit*
0	fric_fric	2(12%)	2(2.6%)	*nerv[ou]s, sax[o]phone*
	fric_stop	10(59%)	56(73.7%)	*pos[i]tive, sens[i]ble*
	r_l	1(6%)	2(2.6%)	*natur[a]lly*
	l_r	1(6%)	54(5.3%)	*gall[e]ry*
	stop_fric	2(12%)	5(6.6%)	*alc[o]hol, opp[o]site*
	stop_stop	1(6%)	7(9.2%)	*t[o]day*
-1	*r*_nasal	2(100%)	2(100%)	*curr[e]ntly, par[a]noid*
-2	*r*_fric	1(10%)	1(3.6%)	*vir[u]s*
	*r*_stop	8(80%)	25(89.3%)	*horr[i]bly, char[a]cter*
	*l*_fric	1(10%)	2(7.1%)	*pol[i]cy*
Total		104	904	

Interestingly, in the cases where the sonority rises by "+2," schwa is more likely to delete between obstruents and [r] than between obstruents and [l]. It can be explained if we distinguish the sonority of [r] and [l], of which the former has higher sonority than the latter (see Selkirk 1984 and Clements 1990, among others). Similarly, in the types where the sonority rises by "+1," we find more cases of schwa deletion between obstruents and nasals than between nasals and liquids/glides. It indicates that although the sonority distance in our analysis is the same, schwa deletion is less likely to occur between two sonorants (e.g. *gen[e]ral, mem[o]ry, fam[i]ly*). This observation is also in accordance to the cases where two consonants are even in sonority; here again, schwa is more likely to delete in sequences involving two obstruents (e.g. *poss[i]ble, alc[o]hol, t[o]day*) than in sequences involving two sonorants (e.g. *gall[e]ry, natur[a]lly*). Finally, in the cases where the sonority falls, schwa deletion is most likely to take place when the schwa is preceded by [r] and followed by a nasal or an obstruent (e.g. *curr[e]ntly, horr[i]ble, par[a]noid*).

4.3 Effect of lexical frequency

Word frequency and word predictability have been suggested in the literature as factors affecting word shortening or reduction. Bybee (2001) proposes that high-frequency words are reduced more often than low-frequency words. Hume (2004) also argues that the more predictable a word is, the less phonetic information is necessary. All these indicate that reduction occurs when words are highly frequent and also highly predictable. The frequency model suggests further that high-frequency words are more likely to be shortened than low-frequency words, whether in the lexicon (Zipf 1929) or during phonetic production (Fidelholtz 1975).

In order to investigate the effect of lexical frequency on schwa deletion, we obtain word frequency from the COCA. The gathered data are then grouped into high and low-frequency words. There are a number of ways by which a threshold between high and low frequency can be established. Bybee (2000) suggests that the high-frequency group has the occurrence of 35 or above per million, and the low-frequency group has less than this number. Patterson *et al*. (2003) indicate that low-frequency words are defined as having a frequency of less than 60 occurrences per million. Coetzee (2010), on the other hand, suggests that one half of the tokens in the corpus are treated as frequent words and the other half as infrequent words. The present study employs the method of Patterson *et al*. in determining the frequency threshold.

The total number of tokens from the COCA is 452,765,36; thus the threshold is set at 27,166 (452,765,365*60/1,000,000). When the frequency of a word from the COCA is 27,166 and over, it is classified as high-frequency words. If its frequency is less than 27,166, it is treated as low-frequency words. Table 8 presents the number of occurrences of schwa presence and deletion in the two frequency categories.

Table 8. Effect of lexical frequency on schwa retention and deletion

	Lexical frequency		Total
	Low frequency	High frequency	
Schwa retention	314(22.9%)	1,059(77.1%)	1,373
Schwa deletion	**141(15.6%)**	**763(84.4%)**	**904**

Table 8 shows that schwa deletion is less frequent for low-frequency words than for high-frequency words (χ^2 = 18.030, p < .000). The overall percentage of schwa deletion in high-frequency words is 84.4% as compared to 15.6% in low-frequency words. The results show that schwa deletion is more common in high-frequency words than in low-frequency words in the BSC (e.g. frequent *average* (89%) and *catholic* (92%) vs. infrequent *alcohol* (21%) and *luxury* (17%)). This finding is consistent with the argument that schwa deletion is most likely to occur in high-frequency words (Hooper 1976, Fokes and Bond 1993, Bybee 2000, Patterson *et al.* 2003).[12]

4.4 Results of the logistic regression analysis

In order to investigate how influential the three linguistic factors are on schwa deletion, we conduct a binary logistic regression. Regression measures the strength of the relationship between dependent and independent variables. In our case, the dependent variable is schwa retention/deletion, and the independent variables are lexical stress environment, sonority distance, and word frequency. Logistic regression is performed because the dependent variable is categorical—whether schwa is absent or present. For the regression analysis, the cases of schwa retention are coded as 0 and those of schwa deletion as 1. The results of the regression analysis are summarized in Table 9.

12 Patterson *et al*. (2003), however, find in the regression analysis that the effect of on the rate of schwa deletion is relatively small. In their analysis based on the Switchboard corpus, pre-stress schwa deletion occurs in 15.4% of high-frequency words and 6.2% of low frequency words.

Table 9. Logistic regression analysis of schwa deletion

	B (SE)	95% C.I for Exp b		
		Lower	Exp *b*	Upper
Stress environment*	**0.531** (0.149)	1.27	1.701	2.277
Sonority distance*	**0.517** (0.058)	1.497	1.677	1.878
Lexical frequency**	**0.268** (0.117)	1.039	1.307	1.645

-2Log likelihood= 2905.918; R^2= 0.065 (Cox & Snell); R^2= 0.088 (Nagelkerke)
Model χ^2=158.378, p < .005. Wald statistic significance: *p < .005, **p < .05

Table 9 reveals that lexical stress environment, sonority distance and lexical frequency are all significant predictors to determine the occurrence of schwa deletion (Wald statistic significance: p < .005 for stress environment and sonority distance; p < .05 for lexical frequency). Of the three linguistic factors, stress environment is the most pervasive factor to account for schwa deletion (this result is consistent with Patterson *et al*. 2003), sonority distance is the next, and lexical frequency is the least influential factor (the regression coefficients are 0.531, 0.517, and 0.268, respectively).

5. Conclusion

The major goals of this paper were to provide a statistical analysis of schwa deletion in American casual speech and to evaluate the effects of the three linguistic factors on schwa deletion. The factors investigated in this analysis were the position of stress within the words, sonority distance between the two consonants surrounding schwa, and word frequency. The findings of this study are as follows.

First, the corpus statistics show that schwa deletion in American casual speech depends on stress position within the word. In general, schwa is more likely to delete in post-stressed position than in pre-stressed position. In three syllable words, post-stress deletion is more common than pre-stress deletion, but we find the opposite distribution in two-syllable words. Second, this study provides clear evidence about the relationship between sonority and schwa deletion. Schwa tends to be deleted as the sonority distance increases between the two consonants that occur before and after schwa. Third, with respect to word frequency, schwa deletion is more frequent in high-frequency words than in low-frequency words. Lastly, the logistic regression analysis shows that all these three linguistic factors play statistically significant roles in the deletion of the unstressed vowel. Of the three linguistic factors, word stress is the most significant predictor of schwa deletion, followed by sonority distance and word frequency.

To conclude, we have shown that stress position, sonority, and word frequency have significant effects on schwa deletion. Nevertheless, we must be careful not to over-interpret these effects because schwa deletion is also conditioned by other linguistic and extra-linguistic factors, which we have not addressed in this paper. We leave the discussion of these factors for future research.

References

Baayen, R. Heral, Richard Piepenbrock, and Leon Gulikers. 1995. *The CELEX Lexical Database (CDROM)*. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Bybee, Joan L. 2000. The phonology of the lexicon: evidence from lexical diffusion. In Michael Barlow, and Suzanne Kemmer (eds.), *Usage Based Models of Language*, 65-86. Stanford: CSLI Publications.

Bybee, Joan L. 2001. *Phonology and language use*. Cambridge University Press.

Clements, George N. 1990. The role of the sonority cycle in core syllabification. In John Kingston, and Mary Beckman (eds.), *Papers in Laboratory Phonology* I, 283-333. Cambridge: Cambridge University Press.

Coetzee, Andries, and Shigeto Kawahara. 2010. Frequency and other biases in phonological variation. *Natural Language and Linguistic Theory.*

Dalby, Jonathan M. 1986. Phonetic structure of fast speech in American English. In *Phonetics and Phonology*. Bloomington, IN: Indiana University Linguistic.

Davidson, Lisa. 2006. Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 64: 79-112.

Davies, Mark. 2010. The Corpus of Contemporary American English: 450 Million Words, 1990-present. Available at http://corpus.byu.edu/coca/.

Fidelholtz, James. 1975. Word frequency and vowel reduction in English. *CLS* 11: 200-213.

Fokes, J., and Bond, Z. S. 1993. The elusive/illusive syllable. *Phonetica* 50: 102-123.

Foley, J. 1972. Rule precursors and phonological change by meta-rule. In Stockwell, R., and R. Macaulay (eds.), *Linguistic Change and Generative Phonology*, 96-100. Bloomington, IN: Indiana University Linguistic.

Glowacka, Dorota. 2001. Unstressed vowel deletion and new consonant clusters in English. *Poznan Studies in Contemporary Linguistics* 37: 71-94.

Godfrey, John J., Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: telephone speech corpus for research and development. *Proceedings of ICASSP*, 517-520.

Hammond, Michael. 1999. Lexical frequency and rhythm. In Michael Darnell, Edith Moravscik, Frederick Newmeyer, Michael Noonan, and Kathleen Wheatley (eds.), *Functionalism and Formalism in Linguistics*, 329-358. John Benjamins.

Hong, Sung-Hoon. 1999. Loss of unstressed vowel in English: prosodic faithfulness through output-output correspondence. *Linguistics in the Morning Calm* 4: 211-226. Linguistic Society of Korea.

Hong, Sung-Hoon. 2007. Variation in English syncope within a partially ranked grammar. *Language* 32(2): 241-263. Linguistic Society of Korea.

Hooper, Joan B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (ed.), *Current Progress in Historical Linguistics*, 96-105. Amsterdam: North Holland.

Hooper, Joan B. 1978. Constraints on schwa-deletion in American English. In Jacek Fisiak (ed.), *Recent Developments in Historical Phonology,* 183-207. The Hague: Norton.

Hume, Elizabeth. 2004. Deconstructing markedness: a predictability-based approach. *Proceedings of the Berkeley Linguistic Society 2004*, 182-198.

Johnson, Keith. 2004. Massive reduction in conversational American English. In K. Yoneyama, and K. Maekawa (eds.), *Spontaneous Speech: Data and Analysis, Proceedings of the 1st Session of the 10th International Symposium.* Tokyo, 29-54. Japan: The National International Institute for Japanese Language.

Kager, René. 1997. Rhythmic vowel deletion in Optimality Theory. In Iggy Roca (ed.), *Derivations and Constraints in Phonology*, 463-499. Oxford: Oxford University Press.

Kiesling, Scott, Laura Dilley, and William D. Raymond. 2006. The Variation in Conversation (ViC) Project: creating of the Buckeye Corpus of conversational speech. Ms. Columbus, OH: Department of Psychology, Ohio State University. Available at http://www.buckeyecorpus.osu.edu.

Ladefoged, Peter. 2006. *A course in phonetics* (5th edition). MA: Thomson Wodsworth.

Patterson, David, Paul C. LoCasto, and Cynthia M. Connie. 2003. Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica* 60: 45-69.

Pérez, Patricia. 1992. Gradient sonority and harmonic foot repair in English syncope. In *Coyote Papers: Working Papers in Linguistic from A to Z* 8: 118-142. University of Arizona.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, Raymond William, Elizabeth

Hume and Eric Fosler-Lussier. 2007. *Buckeye Corpus of Conversational Speech* (2nd release). Columbus, OH: Department of Psychology, Ohio State University (Distributor).

Ryu, Na-Young, and Sung-Hoon Hong. 2011. A corpus-based analysis of schwa deletion in American conversational English. Paper presented at the 2011 Fall Conference, Linguistic Society of Korea.

Ryu, Na-Young, and Sung-Hoon Hong. 2012. The variable deletion of unstressed vowels in American conversational English. Manuscript, Hankuk University of Foreign Studies.

Selkirk, Elizabeth. 1984. On the major class features and syllable theory. In Mark Aronoff, and Richard Oehrle (eds.), *Language Sound Structure*, 107-136. Cambridge: MIT Press.

Song, Jieun. 2013. Variable schwa deletion in English: A corpus study. *Studies in Phonetics, Phonology and Morphology* 19(1): 33-52.

Zipf, George K. 1929. Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40: 1-95.

Zwicky, Arnold M. 1972. Note on a phonological hierarchy in English. In P. S. Robert, and S. M. Ronald (eds.), *Linguistic Change and Generative Theory,* 275-301. Bloomington, IN: Indiana University Press.

Appendix
List of data set from the Buckeye Speech Corpus

A. Word size: 2 is two-syllable words, 3 is three-syllable words
B. Stress environment: 0 is pre-stress position, 1 is post-stress position
C. Sonority distance: -2, -1, 0, 1, 2 (sonority distance indicates the distance between C_1 and C_2 surrounding schwa)
D. Lexical frequency: 0 is low-frequency words, 1 is high-frequency words

No	Word	Word size	Stress environ	Sonority distance	Lexical freq	Num del	Num ret
1	absolute	3	1	2	0	1	2
2	alcohol	3	1	0	0	3	11
3	avenue	3	1	1	0	2	20
4	average	3	1	2	1	16	2
5	bachelor	3	1	2	0	1	0
6	bottom	2	1	1	1	2	12
7	broken	2	1	1	1	2	8
8	buffalo	3	1	2	0	2	2
9	camera	3	1	1	1	3	1
10	career	2	0	2	1	2	12
11	catholic	3	1	2	1	57	5
12	character	3	1	-2	1	3	8
13	collected	3	0	2	0	1	1
14	commit	2	0	1	0	2	5
15	company	3	1	1	1	9	19
16	condom	2	1	1	0	1	3
17	conference	3	1	2	1	1	0
18	connect	2	0	1	0	1	8
19	connected	3	0	1	0	1	5
20	consuming	3	0	1	0	1	1
21	corporate	3	1	2	0	4	0
22	correct	2	0	2	1	2	8
23	correctly	3	0	2	0	1	0
24	correctness	3	0	2	0	1	3
25	coverage	3	1	2	0	1	0
26	cultural	3	1	2	1	1	7

27	currently	3	1	-1	1	1	4
28	dangerous	3	1	2	1	2	8
29	desperate	3	1	2	0	2	0
30	difference	3	1	2	1	38	8
31	different	3	1	2	1	233	128
32	even	2	1	1	1	22	450
33	excellent	3	1	2	0	2	5
34	factory	3	1	2	0	1	3
35	family	3	1	1	1	164	35
36	finally	3	1	1	1	24	7
37	gallery	3	1	0	0	4	4
38	general	3	1	1	1	4	33
39	gorilla	3	0	2	0	1	1
40	grocery	3	1	2	0	1	2
41	happening	3	1	1	0	7	9
42	history	3	1	2	1	27	5
43	horribly	3	1	-2	0	1	4
44	humorous	3	1	1	0	1	0
45	injury	3	1	2	0	3	13
46	interest	3	1	2	1	7	3
47	interests	3	1	2	1	4	1
48	lazarus	3	1	2	0	2	9
49	luxury	3	1	2	0	1	5
50	maverick	3	1	2	0	1	0
51	memory	3	1	1	1	1	0
52	national	3	1	1	0	10	5
53	naturally	3	1	0	0	2	1
54	nervous	2	1	0	0	1	1
55	normally	3	1	1	0	4	7
56	offering	3	1	2	0	1	1
57	opera	3	1	2	0	5	3
58	opposite	3	1	0	0	2	12
59	organize	3	1	1	0	1	1
60	ornery	3	1	1	0	1	0
61	paragraph	3	1	-2	0	4	1
62	parakeet	3	1	-2	0	2	1
63	paranoid	3	1	-1	0	1	1
64	partially	3	1	2	0	1	0
65	personal	3	1	1	1	11	18

66	police	2	0	2	1	11	19
67	policeman	3	0	2	0	1	1
68	policy	3	1	-2	1	2	7
69	positive	3	1	0	1	11	8
70	possible	3	1	0	1	6	8
71	possibly	3	1	0	0	5	7
72	preference	3	1	2	0	1	1
73	principal	3	1	0	0	1	17
74	principle	3	1	0	0	1	4
75	protect	2	0	-2	1	9	4
76	rational	3	1	1	0	1	0
77	reasoning	3	1	1	0	1	0
78	recently	3	1	1	1	1	15
79	reference	3	1	2	0	2	1
80	reverend	3	1	2	0	1	0
81	saxophone	3	1	0	0	1	1
82	sensible	3	1	0	0	1	0
83	separate	3	1	2	1	10	5
84	sesame	3	1	1	0	1	1
85	seventh	2	1	1	0	1	14
86	seventy	3	1	1	0	7	59
87	sickening	3	1	1	0	2	0
88	slippery	3	1	2	0	1	0
89	socially	3	1	2	0	1	5
90	sophomore	3	1	1	0	12	1
91	suffering	3	1	2	0	1	0
92	supply	2	0	0	0	1	0
93	support	2	0	0	1	2	33
94	suppose	2	0	0	0	4	7
95	supposed	2	0	0	1	24	42
96	symphony	3	1	1	0	1	6
97	terrible	3	1	-2	0	2	27
98	therapist	3	1	-2	0	1	2
99	therapy	3	1	-2	0	3	8
100	today	2	0	0	1	7	76
101	tomorrow	3	0	1	1	1	8
102	usually	3	1	2	1	43	52
103	veteran	3	1	2	0	1	2
104	virus	2	1	-2	0	1	0

Na-Young Ryu · Sung-Hoon Hong
Department of English Linguistics
Hankuk University of Foreign Studies
270 Imun-dong, Dongdaemun-gu, Seoul 131-792
E-mail: hongshoon@hufs.ac.kr

Received: 2013. 06. 26
Revised: 2013. 07. 26
Accepted: 2013. 07. 26