# A corpus-based analysis of schwa deletion in American Conversational English

**Na-Young Ryu & Sung-Hoon Hong**

Hankuk University of Foreign Studies

# Table of Contents

# 1.Introduction

- The central purpose of this paper is **to describe the phenomenon of schwa deletion in American conversational English** based on **the Buckeye Speech Corpus(BSC)**.
- We highlight **a number of factors** that play a significant role in schwa deletion and compare the results of the previous studies.
- We offer a **statistical analysis** of the effects of three linguistic factors contributing to variable schwa deletion.

# 2.Previous studies

- **There are numerous factors contributing to schwa deletion.**

(1) Lexical stress environment: pre-stress vs. post-stress (Zwicky 1972, Patterson *et al* 2003))

(2) Word frequency (Fokes and Bond 1993, Hooper 1976)

(3) Sonority relation between the consonants (Zwicky 1972, Hooper 1976, 1978, Pérez 1992, Bybee 2000, 2001)

(4) Speech rate and style (Dalby 1986, Nolan 1992)

(5) Speaker's dialect (Zwicky 1972)

# 3.Potential factors contributing to schwa deletion in this study

- **(1) Lexical Stress environment (pre-stress vs. post-stress)** based on the Buckeye Speech Corpus.

  a. **pre-stress schwa-deletion** occurs before a stressed syllable. (e.g., p[a]ráde, p[o]líce)

4 - 2 = 2

**Police**

C1 C2

- **(3) Sonority condition** that holds between the consonant preceding the target vowel and the consonant following it.

Numerical value: 1 (Glides) 2 (Liquids) 3 (Nasals) 4 (Obstruents)

(e.g., P[o]lice( p_ l):4-2=2)

(clement )

# 4.Buckeye Speech Corpus & COCA

• **The Buckeye Speech Corpus (BSC: Pitt *et al*. 2007)**: consists of 40 individual speaker interviews. The speakers spoke a total of over **300,000 word tokens.** The speech has been orthographically transcribed and phonetically labeled.



• **The Corpus of Contemporary American English (COCA: Davies, 1990~2011)**: contains more than **425 million words of text** and is equally divided among spoken, fiction, popular magazines, newspapers and academic texts. This corpus was used in this study for word frequency.

# 5.How to collect data

1. Collect data from the Buckeye Speech Corpus and the CELEX (Baayen *et al*. 1995)

| The BSC | The CELEX |
|---|---|
| - Total data: 269,648<br>- Information on the orthographic and the phonetic transcription of words | - Total data: 35,804<br>- Information on the number of the syllable and stress environment |

2. Find data which contain both reduced vowels [ɪ, ʌ/ə, ʊ] ([ih,ah,uh]) in the dictionary pronunciation and delete them in the actual pronunciation using perl and excel programs.

| word | Dictionary pronunciation | Actual pronunciation | Schwa deletion (yes/no) | Rate of schwa deletion |
|---|---|---|---|---|
| 'opera' | /aa p ah r ah/<br>/aa p ah r ah/<br>/aa p ah r ah/<br>/aa p ah r ah/<br>/aa p ah r ah/<br>/aa p ah r ah/<br>/aa p ah r ah/<br>/aa p ah r ah/ | [aa p er ah]<br>[aa p er ah]<br>[aa p r ah]<br>[aa p r ah]<br>[ao p er ah]<br>[aw p r ah]<br>[ow p r ah]<br>[ow p r ah] | No<br>No<br>**Yes**<br>**Yes**<br>No<br>**Yes**<br>**Yes**<br>**Yes** | **62.5%** |

# 6.Data

Table 1. Data of schwa deletion from the Buckeye Speech Corpus

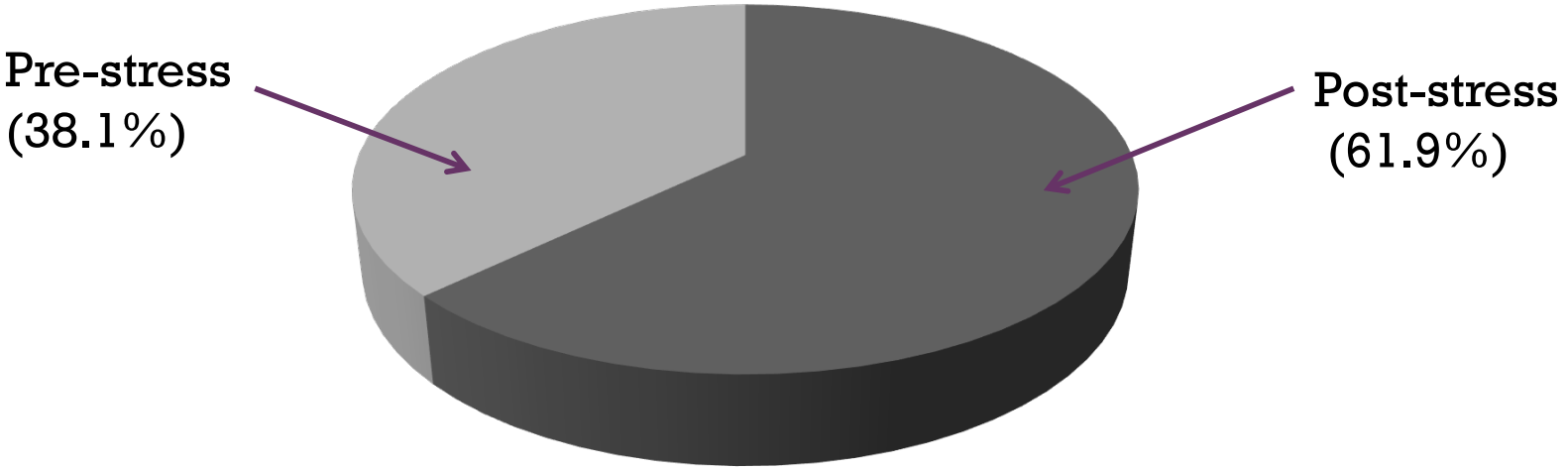
| Total token frequency of schwa deletion | Type frequency of schwa deletion |
|---|---|
| 4,060 | **63** |
| b[e]cause, b[e]lieve, b[e]long, c[o]llect, c[o]mmit, c[o]mmunity, c[o]mputer, c[o]nnect, c[o]rrect, c[o]rrectness, d[e]serve, d[i]vorce, g[a]rage, h[i]larious, ind[i]rectly, p[o]lice, p[o]litical, s[u]pport, s[u]ppose, s[u]pposed, s[u]pposedly, t[o]gether, t[o]ward, t[o]wards<br><br>an[i]mal, av[e]nue, av[e]rage, bas[i]cally, bask[e]tball, cam[e]ra, chath[o]lic, char[a]cter, compass[io]nate, crim[i]nal, cult[u]ral, curr[e]ntly, def[i]cit, def[i]nite, def[i]nitely, diff[e]rent, diff[e]rently, diff[i]cult, disc[i]pline, exc[e]llent, fam[i]ly, fin[a]lly, gall[e]ry, gen[e]ral, gen[e]rally, hist[o]ry, inter[e]st, inter[e]sted, inter[e]sting, nat[io]nal, norm[a]lly, obvi[ou]sly, op[e]ra, pers[o]nal, pers[o]nally, pol[i]cy, pos[i]tive, poss[i]ble, poss[i]bly | |

# Statistical analysis of three factors contributing to schwa deletion

- Frequency analysis
- Multiple regression analysis

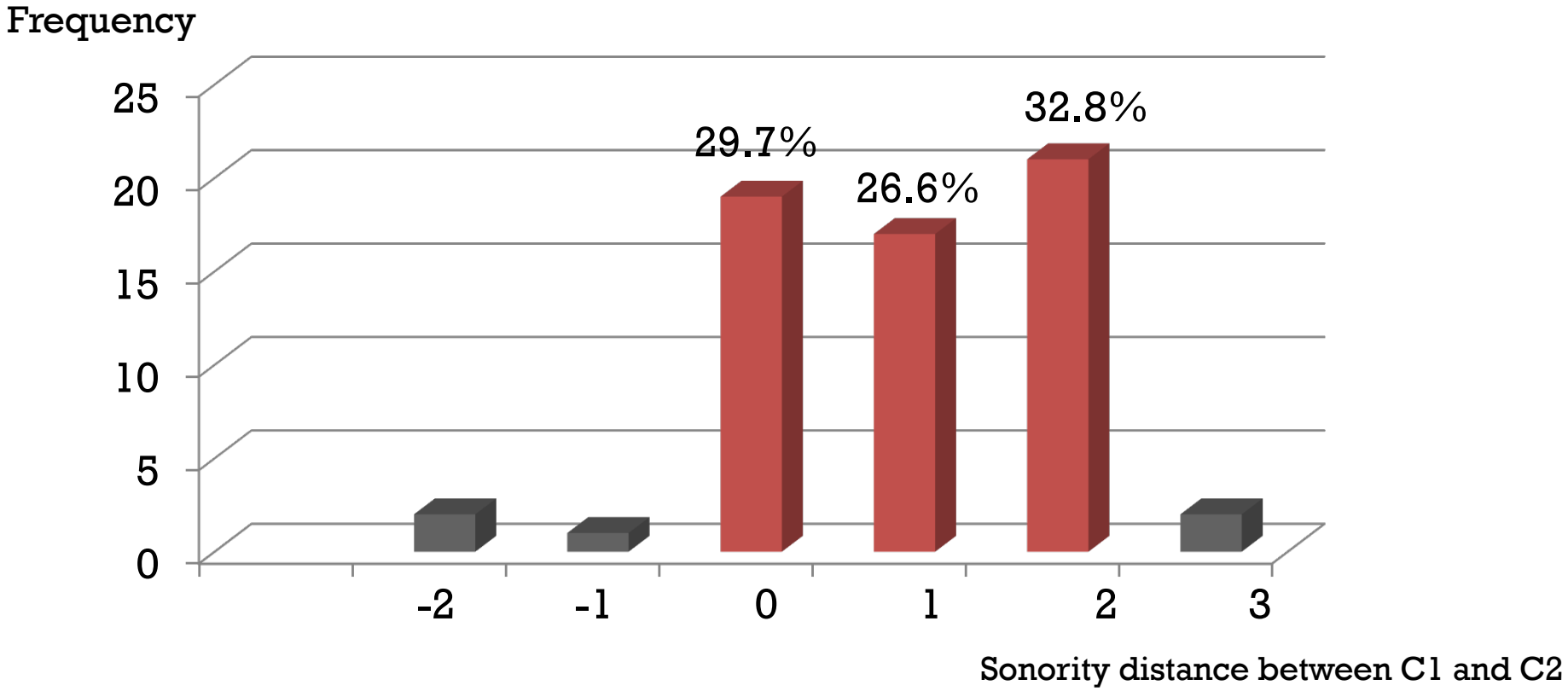
# 7. Effects of lexical stress environment (pre-stress vs. post-stress)

| | Pre-stress | Post-stress |
|---|---|---|
| Type Frequency | 24 | 39 |
| Example | c[o]'llect, p[o]'lice | 'an[i]mal, 'cam[e]ra |




A schwa is more likely to appear in post-stress environment than in pre-stress environment.

# 8. Effects of sonority difference between C1 and C2




pol[i]cy curr[e]ntly b[e]cause c[o]mputer c[o]rrect t[o]ward

# 11.Multiple regression analysis of schwa deletion

$R^2 = .218$

| Dependent variable | Independent variables |
| --- | --- |
| The rate of schwa deletion (from BSC) | 1.Stress environment (p=.001) (pre-stress vs. post-stress) |
| | 2.Sonority distance between C1 and C2 (p=.012) |
| | 3.Word frequency (p=.953) (from COCA) |

# 11.Discussion

- The results reveal that the stress environment (pre-stress vs. post-stress, p-value=.001) and the sonority distance between C1 and C2(p-value=.012) are statistically significant in schwa deletion.

- For future research, we will include more factors such as the number of syllables, speech rate and style in order to obtain more in-depth results of the phenomenon of schwa deletion in American conversational English.

# References


- Bybee, Joan L. (2000). The phonology of the lexicon: evidence from lexical diffusion.
- Bybee, Joan L. (2001). *Phonology and Language Use*.
- Dalby, Jonathan M. (1998). Phonetic structure of fast speech in American English
- Fokes, Joann. & Bond, Z.S. (1993). The elusive/illusive syllable.
- Hooper, Joan B. (1978). Constraints on schwa-deletion in American English.
- Pérez, Particia (1992). Gradient Sonority and Harmonic Foot Repair in English Syncope.
- Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. and Fosler-Lussier, E. (2007) Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu]
- Zwicky, Arnold ( 1972). Note on a Phonological Hierarchy in English.