THE EFFECT OF ALLOPHONIC PATTERNS ON CONSONANT CLUSTER SIMPLIFICATION IN AMERICAN ENGLISH

Caroline L. Smith

University of New Mexico
caroline@unm.edu

ABSTRACT

Previous work on consonant cluster simplification in English has shown that a consonant surrounded by other consonants is more likely to delete than a consonant adjacent to a vowel. A consequence is that in three-consonant word-final clusters, the penult is most likely to delete. Experimental data also shows more frequent deletion of word-final stops in words that frequently occur before consonant-initial words [16].

We look at a related issue using conversational speech (Buckeye Corpus, [12]): if speakers frequently encounter a cluster produced with an epenthetic consonant resulting from an allophonic process, does that influence their production of similar clusters? Higher rates of epenthesis in /-ns#/ clusters are mirrored by low rates of /t/ deletion in /-nts#/, compared to low epenthesis in /-nz#/ but high deletion in /-ndz#/. These high-frequency clusters show more effect of contextual frequency than do lower-frequency comparable clusters with /l/.

Keywords: consonant clusters; frequency; consonant deletion; conversational speech.

1. INTRODUCTION

Many factors that influence the frequency of consonant cluster simplification (the deletion of one or more consonants in a cluster) have been studied in English and other languages. Previous work (e.g. [17, 18, 21]) has shown that following context is a powerful factor: when a consonant is followed by another consonant, it is more likely to delete than when it is followed by a vowel.

Less studied than immediate segmental context is the cumulative effect of the different contexts where a word or sound occurs, which in a usage-based approach is expected to affect speakers' productions. Brown, Raymond, and colleagues [15, 16] have found that frequent occurrence of a word or sound in a context that favors a variable phonological process increases the overall likelihood of that process occurring, even in tokens that occur in less favorable contexts. Raymond and Brown [15] refer to this factor as "Frequency in a Favorable Context" (FFC). Their analyses show that FFC has a stronger effect than the raw frequency of an individual word. The argument is that language users' mental representation of a word is modified by their experiences of that word: if they frequently encounter it in a reduced form, because it tends to occur in environments where reduction is likely, then the reduced form strengthens in the speaker's mental lexicon. Raymond et al. [16] use this idea to show that frequent occurrence in consonantal contexts increases the likelihood of English word-final t/d deletion in experimental data. They found that although the context surrounding a word varies from token to token, speakers are influenced by the frequency with which they have encountered a sound in a context favorable to deletion.

Raymond and Brown's work looks at cases where the context of a word varies depending on which words surround it, that is, which bigrams it occurs in. In this paper, we extend the concept of FFC by examining the effect of contextual variability that arises from an optional allophonic pattern. If speakers frequently encounter a particular sequence of sounds, that could bias them towards applying, or not applying, an optional phonological process so as to match the frequently-occurring sequence. Here the sequences studied are consonant clusters. The two processes under consideration have opposite effects: one is the simplification of consonant clusters by deletion of one member, and the other is allophonic epenthesis of oral stops in sonorant-fricative clusters.

1.1. Consonant cluster simplification

Although English allows up to four consonants at the end of a syllable, the absence of one or more of these consonants is common. In this paper the term "deletion" is used, although it has been shown (e.g. [1, 3]) that re-timing of the gestures for a sequence of consonants can "hide" other consonants that are in fact articulated. Numerous studies of consonant deletion in clusters show that local segmental context influences the likelihood of deletion (e.g. [7, 18, 22]). A preliminary analysis [21] of three-consonant word-final clusters in the Buckeye Corpus showed that the penultimate consonant in the cluster was the most likely to be deleted, and this likelihood increased when the next word began with a

consonant, even though that consonant was not immediately adjacent.

1.2. Sonorant-fricative epenthesis

Epenthesis in sonorant-fricative clusters is one of the most described allophonic patterns in American English. It is most often described in nasal-fricative clusters [2, 10, 20, 24], but some authors also discuss similar patterns in lateral-fricative clusters [5, 6, 11, 19]. Plath [14] writing in 1958, claimed that /-n[t]s/ and /-nts/ were homophonous in his speech, but that he contrasted /-ls/ and /-lts/. Here we are not concerned with the specifics of how the epenthetic /t/s are produced or perceived, but with the contexts in which they occur. Although [10] argue that this change (stop epenthesis) has almost reached completion, resulting in nearly complete neutralization, the results reported here suggest that it is not yet so regular, at least in the dialect and speech style represented in the Buckeye Corpus.

To the best of my knowledge, [2], also discussed in [24], is the only previous study to examine the rate of production of epenthetic stops in nasal-fricative clusters in a large corpus of American English. Blankenship analyzed these clusters in the TIMIT corpus, which consists of read sentences rather than conversational speech as used for the present study. She observed epenthetic [t] in 26% of /-ns/ clusters, and a loss of /t/ in 14% of /-nts/ clusters. Experimental studies (e.g. [6, 20]) have observed rates of epenthesis ranging from 71-82% [20] to 100% [6] in /-ns/ clusters.

A suggestion that frequency plays a role in epenthesis comes from [10]'s finding that higher word frequency improved discrimination between epenthetic and underlying stops: listeners were better able to distinguish higher frequency words with epenthetic stops from their counterpart with an underlying stop (*prince* vs. *prints*), compared to lower frequency pairs which were more difficult. They suggest that this is due to a less robust mental representation of the lower-frequency words.

2. METHOD

This paper is concerned with word-final clusters of two or three consonants, in which the first consonant is a sonorant (either /n/ or /l/), the second, when present, is an alveolar stop (/t/ or /d/), and the last is a fricative (/s/ or /z/).

The Buckeye Corpus was used as a source of conversational speech [12, 13]. It consists of recordings of 40 speakers of English from the Columbus, Ohio, area, with equal numbers of male and female, younger and older speakers. It provides both word-level orthographic transcription and phonetic transcription aligned with the acoustic waveform and segmented at the phone level. Previous investigation of consonant deletion in all three-consonant word-final clusters in the Buckeye Corpus [21] showed that while /-nts/ is the most frequent cluster, rates of stop deletion are higher in /-ndz/, which is the second most frequent cluster. This finding prompted more study of these clusters.

An exhaustive search of the corpus was conducted using Phonological Corpus Tools [8] to identify all words ending in the two-consonant clusters /-ns/, /-nz/, /-ls/, and /-lz/, and the three-consonant clusters /-nts/, /-ndz/, /-lds/, and /-ldz/. The search for words excluded those transcribed with syllabic [n] or [l], which the Buckeye transcription alphabet distinguishes from the purely consonantal forms of these sounds. The `.words` files from the corpus (which list the orthographic form and time stamp of each word) were then searched to identify all occurrences of the identified words. Transcriptions of the citation pronunciation and the actual pronunciation of each token were also extracted from the `.words` files.

2.1. Transcription issues

The Buckeye Corpus lists a citation pronunciation of each word. These were used as is except for a few minor corrections of what appeared to be errors ([z] replaced erroneous [s] in words ending *–ands*). The corpus transcriptions of actual pronunciations provide more phonetic detail than the citation pronunciations. They distinguish nasalized vowels from oral ones. For the present study, if a cluster's canonical pronunciation includes a nasal consonant, the actual pronunciations were classified as canonical if they included a nasal consonant or a nasalized vowel in the appropriate position.

A number of the actual pronunciations showed palatalization of the final fricative, where [ʃ] appeared instead of [s], or [ʒ] instead of [z]. These almost always occurred preceding a word beginning with [j]. Palatalizations of this form were ignored in deciding whether to classify a pronunciation as canonical. Rare cases where the final consonant was transcribed as [tʃ] were classified as including an oral stop. There were no final [dʒ].

3. RESULTS

3.1. Nature of the data

Table 1 shows the total numbers of tokens and types (distinct words) containing each of the clusters. The /-ls/ cluster has by far the highest ratio of tokens to types, because the data are dominated by 174 occurrences of *else* (94% of the tokens of that

cluster). Similarly, the data for /-lts/ are dominated by 16 occurrences of *adults* (84% of the tokens). For /-nts/, there are 198 tokens of *parents* (36% of the total); for /-ndz/, there are 134 tokens of *friends* (38% of the total). For other clusters, no one word provides a large proportion of the data.

Table 1: Number of tokens for each of the clusters. Token frequency is the number of occurrences; type frequency is the number of different words containing that cluster.

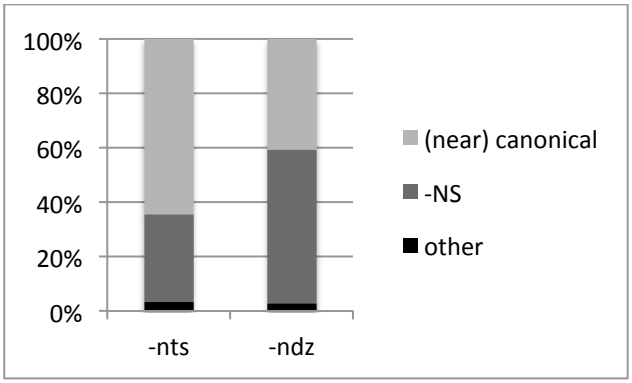
Word	Token Frequency	Type Frequency
-ns	660	94
-nz	986	246
-nts	543	77
-ndz	352	43
-ls	185	12
-lz	400	69
-lts	19	4
-ldz	22	8

3.2. Actual pronunciations: nasal clusters

3.2.1. Cluster simplification

As expected, deletion of the cluster-medial /t/ or /d/ was very frequent, especially in the voiced clusters. The difference between /-nts/ and /-ndz/ is significant at $p<.001$ ($\chi^2 = 52.15$). For comparison, [17] report a deletion rate of 16.5% of /t/ and /d/ in word-medial clusters in the Buckeye Corpus.

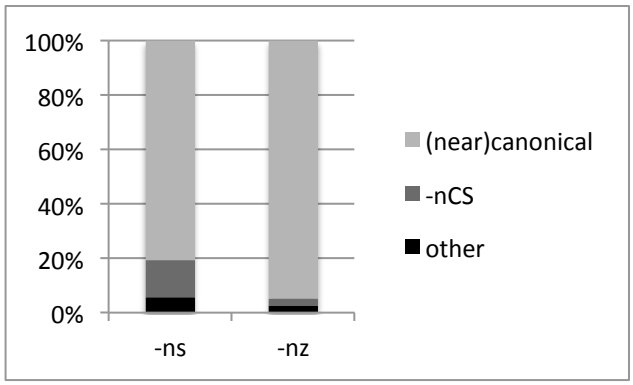
Figure 1: The percentage of tokens produced with canonical or near-canonical pronunciations, with deletion of the oral stop (-NS), or in other forms, for words with word-final /-nts/ and /-ndz/ clusters.



3.2.2. Epenthesis

The overall rate of epenthesis in /-ns/ clusters was 13%, considerably lower than what is reported by other studies. The percentages of different pronunciations are plotted in Figure 2, for the /-ns/ and /-nz/ clusters. The difference in epenthesis rate between the voiceless and voiced clusters was significant at $p<.001$ ($\chi^2 = 65.11$).

Figure 2: The percentage of tokens produced with canonical or near-canonical pronunciations, with an epenthetic stop, or in other forms, for words with word-final nasal-fricative clusters.



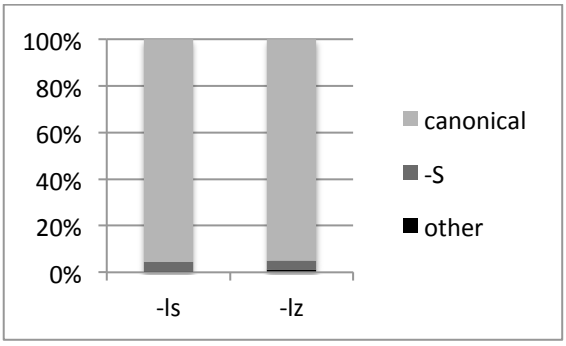
The absence of epenthesis in /-nz/ clusters is consistent with the frequent deletion of [z] in the /-ndz/ clusters: speakers expect these voiced clusters to occur without a medial stop. Similarly, the more frequent epenthesis in /-ns/ is consistent with the lower deletion rate of [t] in /-nts/.

3.3. Actual pronunciations: lateral clusters

The lateral clusters were far less variable than the nasal clusters. There were very few tokens of /-lts/ and /-ldz/, but even so the lack of variability is striking: 17 of 19 /-lts/ clusters and all /-ldz/ clusters were produced with the canonical pronunciation.

The two-consonant lateral clusters show only a little more variation. Epenthesis was found in one token of /-ls/ and never in /-lz/. More common was deletion of the lateral, in 7 /-ls/ and 15 /-lz/, unlike the nasal clusters where the absence of any nasal was less common. These are plotted in Figure 3.

Figure 3: The percentage of tokens produced with canonical or near-canonical pronunciations, with deletion of the lateral (-S), or in other forms, for words with word-final lateral-fricative clusters.



3.4. Rates of epenthesis in /-ns/ clusters

For individual words ending in /-ns/, the percentage of tokens produced with epenthetic stops ranged from 0% to 100%. For words with very few tokens, these percentages are not informative. Words with more than 10 tokens are examined in more detail here, and listed in Table 2.

Table 2: Proportion of tokens with epenthesis for words ending in /-ns/ for which there were more than 10 tokens available.

Word	n	Epenthesis Rate
once	110	9%
since	87	8%
difference	46	17%
sense	33	39%
experience	36	14%
violence	30	37%
influence	29	17%
science	20	5%
chance	20	10%
insurance	17	0%

Of these, the epenthesis rates for *insurance* (0%) and *science* (5%) were well below the overall average, while the rates for *sense* (39%) and *violence* (37%) were the highest. The only salient phonological difference among these words is that *sense* is monosyllabic, with the cluster in a stressed syllable, while the other words are polysyllabic, and the cluster is in an unstressed syllable. Since the cluster in *violence* with a high rate of epenthesis occurs in the same environment as the clusters in *science* and *insurance*, these differences seem unlikely to be explanatory. ([24] also found that lexical stress did not predict epenthesis rates.)

The lexical contexts in which these four words occur were examined. The bigram frequencies for each target word in combination with every word that preceded or followed it in the Corpus were determined, and the overall means for these bigram contexts were calculated, shown in Table 3. A bigram frequency could not be determined for every token, because the target word was often preceded or followed by silence or speech from the interviewer.

Insurance and *science*, which had very low rates of epenthesis, have lower bigram frequencies than *sense* and *violence*, with high rates of epenthesis. "Frequency in a Favorable Context" offers a possible interpretation of this result: epenthesis is more acceptable to speakers in the higher-frequency contexts because they are likely to have previously encountered the words in these same contexts. Higher bigram frequency has been shown to increase the occurrence of other connected-speech processes such as inter-word palatalization [4]. This result is also consistent with Shosted's [20] finding that speakers produce fewer epenthetic stops in focus conditions: words (such as *insurance* and *science*) that occur in less frequent contexts are more likely to be salient to the speaker, akin to being under focus.

Table 3: Of the words ending in /-ns/ with more than 10 tokens, the mean preceding and following bigram frequencies for the two words with the lowest and highest rates of epenthesis, averaged across all tokens of the words.

Word	Preceding Bigram Freq.	Following Bigram Freq.
insurance	2.6	1.3
science	1.9	2.1
sense	3.9	3.8
violence	3.4	4.3

4. DISCUSSION

This paper argues that speakers' experience of variant pronunciations of consonant clusters reinforces their production of similar clusters, consistent with related research showing that the production and perception of variant pronunciations can be influenced by the distribution of experience of a word in different contexts [16, 23]. These results are consistent with a model where speakers store variant pronunciations in the lexicon, including potentially forms with deletion of a segment [9]. Repetition of related forms reinforces their representations, increasing the likelihood that future productions will resemble the variants most often encountered.

One surprising finding here was the relatively low rates of epenthesis in /-ns/ clusters. A possible explanation is that in the process of transcribing the Buckeye Corpus [13], the criteria for justifying an epenthetic segment were relatively stringent, either deliberately or because of conservative behavior by the transcribers. In studies such as [2], [6], [20], the acoustic signals were evaluated specifically with respect to epenthetic stops, and this may have led to more liberal transcriptions. Another possibility is that an experimental setting encourages more epenthetic stops, possibly as a characteristic of clear speech. ([2] is a corpus study, but using individual sentences read in a controlled environment.) Further investigation is needed to understand the consequences of different contexts and styles on this and other variable phonological processes.

5. REFERENCES

[1] Barry, M. 1985. A palatographic study of connected speech processes. *Cambridge Pap. in Phonetics and Expt'l Linguistics* 4, 1-16. Dept. Ling., Cambridge.

[2] Blankenship, B. 1992. What TIMIT can tell us about epenthesis. *UCLA Work. Pap. in Phon.* 81, 17-25.

[3] Browman, C., Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. In: Kingston, J., Beckman, M. (eds), *Pap. in Lab. Phonol.* I. Cambridge: Cambridge University Press, 341-375.

[4] Bush, N. 2001. Frequency effects and word-boundary palatalization in English. In: Bybee, J., Hopper, P. (eds), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins, 255-280.

[5] Clements, G.N. 1987. Phonological feature representation and the description of intrusive stops. In: Bosch, A., Need, B., Schiller, E. (eds), *CLS 23: Parasession on Autosegmental and Metrical Phonology*, 29-50.

[6] Fourakis, M., Port, R. 1986. Stop epenthesis in English. *J. Phon.* 14, 197-221.

[7] Guy, G. 1980. Variation in the group and the individual: the case of final stop deletion. In Labov, W. (ed), *Locating Language in Time and Space*. London: Academic Press, 1-36.

[8] Hall, K. C., Allen, B., Fry, M., Johnson, K., Lo, R., Mackie, S., McAuliffe, M. 2017. Phonological CorpusTools, Version 1.3. [Computer program]. Available from PCT GitHub page.

[9] Hanique, I., Ernestus, M., Schuppler, B. 2013. Informal speech processes can be categorical in nature, even if they affect many different words. *J. Acoust. Soc. Am*. 133, 1644-1655.

[10] Kilpatrick, C., Shosted, R., Arvaniti, A. 2007. On the perception of incomplete neutralization. *Proc. 16th ICPhS* Saarbrücken, 653-656.

[11] Ohala, J.J. 1974. Experimental historical phonology. In: Anderson, J., Jones, C. (eds), *Historical linguistics II. Theory and description in phonology*. Amsterdam: North Holland, 353-389.

[12] Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., Fosler-Lussier, E. 2007. Buckeye Corpus of Conversational Speech (2nd release). http://www.buckeyecorpus.osu.edu Columbus, OH: Department of Psychology, Ohio State University (Distributor).

[13] Pitt, M., Johnson, K., Hume, E., Kiesling, S., Raymond, W. 2005. The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Comm.* 45, 89-95.

[14] Plath, W. 1958. The relative frequency of English consonantal patterns. *Zeitschrift für Allgemeine Sprachwissenschaft* 11, 67-87.

[15] Raymond, W., Brown, E. 2012. Are effects of word frequency effects of context of use? An analysis of initial fricative reduction in Spanish. In: Gries, S., Divjak, D. (eds), *Frequency Effects in Language Learning and Processing*. Berlin: de Gruyter, 35-52.

[16] Raymond, W., Brown, E., Healey, A. 2016. Cumulative context effects and variant lexical representations: Word use and English final t/d deletion. *Lang. Var. and Change* 28, 175-202.

[17] Raymond, W., Dautricourt, R., Hume, E. 2006. Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical and phonological factors. *Lang. Var. and Change* 18, 55-97.

[18] Schuppler, B., van Dommelen, W., Koreman, J., Ernestus, M. 2012. How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *J. Phon.* 40, 595-607.

[19] Shinya, T. 2005. The perception of epenthetic stops in English: The effects of cluster type and silent interval. *U. Mass. Occasional Papers in Ling.* 31: *Papers in Expt'l Phonetics and Phonology*, 121-150.

[20] Shosted, R. 2011. An articulatory-aerodynamic approach to stop excrescence. *J. Phon.* 39, 660-667.

[21] Smith, C.L., Amador, B., Brown, H., Cho, G., Fernandez, M., Hobbs, D., Martinez, B., Philp, L., Van Gysel, J.E.L., Yazzie, T. 2018. Patterns of word-final consonant cluster simplification in American English conversational speech. 13th High Desert Ling. Soc. Conference, Nov., https://hdls.unm.edu/assets/images/hdls13-schedule/hdls13_abstracts.pdf.

[22] Tagliamonte, S., Temple, R. 2005. New perspectives on an ol' variable: (t,d) in British English. *Lang. Var. and Change* 17, 281-302.

[23] Viebahn, M., Luce, P. 2018. Increased exposure and phonetic context help listeners recognize allophonic variants. *Attention, Perception, & Psychophysics* 80, 1539-1558.

[24] Yoo, I., Blankenship, B. 2003. Duration of epenthetic [t] in polysyllabic American English words. *J. Int'l Phonetic Assoc*. 33, 153-164.