The Role of the Lemma in Form Variation

Daniel Jurafsky, Alan Bell, and Cynthia Girand

University of Colorado, Boulder

November 18, 2000

1 Introduction

A key problem in building a complete model of the lexicon is understanding the complex relationship between semantically and syntactically defined lexical entries ('lemmas' in the terminology of Levelt (1983)), and phonological forms ('word-forms' or 'lexemes'). One reason for the complexity is that the relationship is not one-to-one. For example in homophones like *still*, a single wordform /stɪl/ is linked with many different lemmas, including verbs ('to quiet'), nouns ('equipment for distilling' or 'silence'), adverbs ('yet' or 'nevertheless'), and adjectives ('silent', 'not moving'). The complementary situation of a single lemma linked with multiple wordforms includes well-known instances of allomorphy such as the realization of the English indefinite article as *a* or *an*, or the definite article *the* as [ðə] or [ði].

In addition to such lexically based form variation, recent corpus-based studies have shown extensive form variation in the pronunciations of what may be single lemmas (Jurafsky et al., 1998; Bell et al., 1999). For example, the phonetically transcribed Switchboard corpus (Greenberg et al., 1996) contains 33 pronunciations for the word *the*. Besides the two allomorphs [ðə] and [ði], most of the other surface realizations are due to segmental and prosodic context and to production factors like speaking rate. But in some cases, where multiple lemmas are mutually associated to a phonological form, it may be that the lemma difference plays a role in the surface realization. The impressionistic summary of Roach (1983), while noting the extensive contextual variation for most English function words, describes lemma-based variation for different functions of *that*, *some*, *there*, and *must*. Such a difference has been repeatedly noted between the pronoun *that* (*That was pretty heart-rending for her*) and the complementizer *that* (*I can't say that I'm an expert on the region*) (Jones, 1947; Jespersen, 1933; Berkenfield, 2000).

In this paper we examine the role of the lemma in explaining these kinds of variation, focusing on four very frequent function words; *to*, *that*, *of*, and *you*. The variation in these four function words provides a suitable locus for this study for several reasons. First of all, they have distinguishable functions that are plausibly instantiated as different lemmas. Second, the functions are plausibly assumed to share phonological forms or lexemes. Third, they exhibit extensive surface variation. Fourth, they are very frequent, permitting enough observations to distinguish the various sources of the variation. For example, the word *to* is very frequent (occurring 68,352 times out of about 3 million words in Switchboard, thus consisting of about 2% of the word tokens), is commonly assumed to have a single lexeme/wordform /tu/, has (at least) two lemmas—an infinitive marker (*we had* **to** *do it*) and a preposition (*I would have gone* **to** *the store*), and has a lot of surface variation (appearing as [tu], [tə], [tɨ], [tʌ], etc.).

We also chose to investigate function words, rather than content words, because function words are much less likely to receive sentence accent, which interacts strongly with reduction, and which we could not control adequately in our corpus. Note that the term *function word* is here no more than a convenient descriptive label; our study will not allow us to make any claims about, for example, differential processing of function and content words, or the exact nature of the set of function words.

When two or more lemmas systematically vary in their surface realizations there are three main classes of explanations for the variation:

1. **Contextual:** The variation is due to contextual and production factors acting on a single phonological form, as sketched in Figure 1.

2. **Multiple lexeme:** The variation is due to multiple phonological forms, possibly differentially linked to the lemmas (Figure 2).

3. **Lemma-based:** The variation is due to differences in the lemmas (e.g. frequency) sharing a single phonological form (Figure 3).

These explanations are of course not exclusive, and many combinations and variants are possible for any given case.

The contextual explanation, sketched in Figure 1, suggests that pronunciation variation is not represented in any way in the lexicon. Rather, whatever variation in wordform we see depends only on contextual factors such as prosodic, segmental or syntactic context, production factors like rate of speech, or sociolinguistic factors. This explanation of variation appears to be the cause of most kinds of variation, if impressionistic summaries like Roach's or the relative rarity of diachronic homonym splits like that of New York and Philadelphia *can* are any guide (Ferguson, 1975; Labov, 1989).

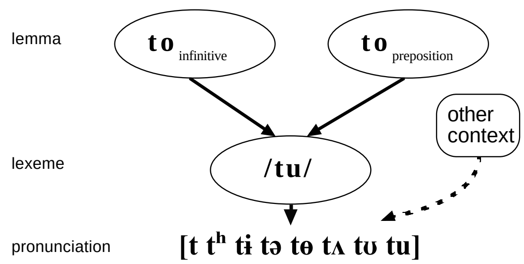

Figure 1: The contextual model: two lemmas share a wordform but there is no effect of these different lemmas on wordform variation. Variation is accounted for solely by effects of context not mediated through the lemma or through the wordform.

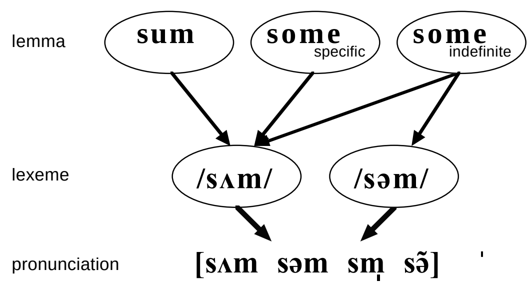

Figure 2: The multiple lexeme model. Different lemmas are linked in overlapping ways to different wordforms.

An example of the multiple lexeme explanation is given by the various senses of the word *some* (and the homonym *sum*), shown in Figure 2. Roach (1983) points out that there are pronunciation differences between two senses of *some*, the specific *some animal broke it* and the indefinite *have some more tea*. These differences are most directly accounted for by assuming that indefinite *some* is linked to two wordforms /səm/ and /sʌm/, while specific *some* is linked only to /sʌm/. (*Sum* is also of course linked only to /sʌm/.) Figure 2 sketches this kind of lemma effect.

Similarly, although the split between the New York and Philadelphia English noun *can* (*can of beans*; /keːən/) and auxiliary *can* (*can I*; /kæn/) may have had its original source in a prosodically conditioned longer noun *can* (perhaps also abetted by its lower lexical frequency), it likely passed through a stage where the two lemmas differentially selected multiple wordforms.

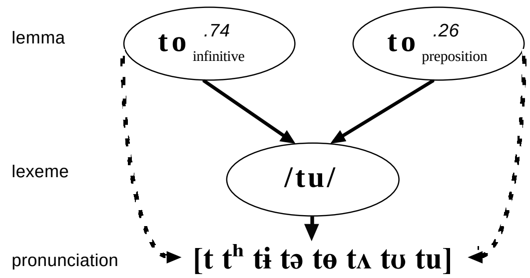

Figure 3: The lemma-based model: Two lemmas share a wordform and some of the wordform variation is accounted for by the effects of lemma differences (such as frequency) on lexical access and/or other production processes. The lemmas *too* and *too* are omitted.

The lemma-based explanation, sketched in Figure 3, assumes that the variation in surface form is somehow caused directly by a property of the different lemmas, such as their frequencies. For example, if the lemmas differ in frequency, this difference might account for some of the surface variation, even though they share a single phonological form. One way that this might happen is for less frequent lemmas, being accessed more slowly, to sometimes slow up one or more of the steps in the phonetic encoding process. We will call this the lemma frequency hypothesis.

The lemma frequency hypothesis is not compatible with current phonological theory and most models of speech production, which generally assume that multiple lemmas have no direct effect on surface variation (Levelt et al., 1999). Furthermore, previous research such as Jescheniak and Levelt (1994) has argued convincingly that the wordform and not the lemma is the major locus of frequency effects in the lexicon, at least so far as they affect lexical access speed in production. Still, we feel it is worth examining the lemma frequency effect. One main reason is that there is robust evidence for a lemma frequency effect in comprehension. For example Hogaboam and Perfetti (1975), Simpson (1981), Simpson and Burgess (1985), and many more recent studies have found that the rate of rise in activation for a particular lemma in comprehension is a function of the lemma's relative frequency. Since there are many differences between lexical comprehension and production, it is crucial to understand exactly which kinds of frequency effects are not symmetric in this way.

A second reason to investigate the lemma frequency effect is that an initial, raw measurement of word durations does seem to give preliminary evidence of a frequency effect. For example, as we will show below, in our corpus the more frequent infinitive lemma for *to* is much shorter on average (with an average raw length of 110 ms) than the less frequent preposition lemma for *to* (with an average raw length of 140 ms). Berkenfield (2000) found similar relations between frequency and shorter duration for the word *that* in a different speech corpus.

Lemma frequency is also a convenient avatar for general lemma effects because in addition to predicting that the lemma will affect the reduction of the wordform, it makes a more specific prediction about the direction of the reduction: more frequent lemmas will have shorter wordforms.

2 Methodology

This paper is based on a new corpus-based methodology to explore the interaction of frequency, lemma, and wordform. Our method is to show how various frequency factors affect the duration, reduction, or lenition of words in a corpus. That is, rather than using carefully controlled materials in production studies in laboratory settings, we use a very large database of natural speech and use multiple regression to control for factors that influence the duration or reduction of words. Our study is based on the Switchboard corpus of 2430 telephone conversations between strangers, collected in the early 1990s (Godfrey et al., 1992).

We coded each instance of the four words *to*, *that*, *of*, and *you* for their different syntacto-semantic lemmas. For example, we investigated four different *that*s: (complementizer, pronoun, determiner, and relative pronoun) and two different *to*s (preposition, infinitive marker). Lists and examples of each sense are given below. The different lemmas of a word may differ in frequency. The infinitive *to*, for example, is about 3 times as frequent as the preposition *to*.

The phonological variable we used to study these parts of speech was the reduction or lenition of the word's pronunciation in conversational speech. Based on our previous work (Jurafsky et al., 1998; Bell et al., 1999) we coded three measures of reduction for *that*, *to*, *of*, and *you*: duration in milliseconds, reduced vowel, and deletion of final consonants

In earlier work (Jurafsky et al., 2001), we have shown that this methodology is quite sensitive to word frequency. In natural speech, frequent words are shorter in duration, have a greater proportion of reduced vowels, and are more likely to have deleted coda consonants than rarer words. This effect holds even after controlling for the many factors that we and others have shown influence reduction in function words, including the speaker's rate of speech (in syllables per second), whether the speaker was having planning problems (as indicated by neighboring disfluencies), the position of the function word in the utterance, the segmental context, the contextual predictability of the function word, and sociolinguistic factors such as age and sex.

In the experiments described in this paper, then, we controlled all our data for these factors and then tested whether the more frequent lemma (e.g., the infinitive *to*) was shorter than the less frequent lemma (e.g., the preposition *to*). Since we did see such a difference in raw durations (for example infinitive *to* is in fact shorter on average than preposition *to*) we expected to see this difference after controlling for other factors.

2.1 The Switchboard dataset

Our dataset of the four function words was drawn from the Switchboard corpus of telephone conversations between strangers, collected in the early 1990s (Godfrey et al., 1992). The corpus contains 2430 conversations averaging 6 minutes each, totaling 240 hours of speech and about 3 million words spoken by over 500 speakers. The corpus was collected at Texas Instruments, mostly by soliciting paid volunteers who were connected to other volunteers via a robot telephone operator. Conversations were then transcribed by court reporters into a word-by-word text.

Approximately four hours of speech from these conversations were phonetically hand-transcribed by students at UC Berkeley (Greenberg et al., 1996) as follows. The speech files were automatically segmented into pseudo-utterances at turn boundaries or at silences of 500 ms or more, and a rough automatic phonetic transcription was generated. The transcribers were given these utterances along with the text and rough phonetic transcriptions. They then corrected the phonetic transcription, using an augmented version of the ARPAbet, and marked syllable boundaries, from which durations of each syllable were computed.

This phonetically transcribed corpus, contains roughly 38,000 transcribed words (tokens). Our total dataset is drawn from the set of all instances of *of*, *that*, *you*, and *to*, after screening for transcription errors. Each of our analyses of reduction are based on the tokens remaining after excluding various non-comparable items, as explained below in section 2.3.

Each observation was coded for two or three factors reflecting reduction:

vowel reduction: We coded the vowel of each function word as *full* or *reduced*. The full vowels included basic citation or clarification pronunciations, e.g. [ði] for *the*, as well as other non-reduced vowels. The reduced vowels that occurred in the function words were [ə] and [ɨ].[1] Table 1 shows full and reduced-vowel pronunciations of the four words.

duration: the duration of the word in milliseconds.

[1]In general we relied on Berkeley transcriptions for our coding, although we did do some data cleanup, including eliminating some observations we judged likely to be in error; see Jurafsky et al. (1998) for details.

coda obstruent deletion: for *that* and *of*, whether the word-final obstruent was deleted.

	Full	**Reduced**
of	[ʌv],[ʌ],[ʌvv] [ɪ],[i],[ɑ]	[ə],[əv],[əf]
to	[tu],[ɾu] [tʊ],[tɪ],[tʌ]	[tə],[tɨ],[ə]
that	[ðæ],[ðæt], [æ] [ðɛ],[ðɛt],[ðɛɾ]	[ðɨt], [ðɨ], [ðɨɾ]
you	[yu],[u], [yɪ],[ɪ],[i]	[yɨ],[y],[ɨ]

Table 1: Common pronunciations of the four function words by vowel type.

2.2 The regression analysis

We used multiple regression to evaluate the effects of our predictability factors on reduction. A regression analysis is a statistical model that predicts a *response variable* (in this case, the word duration, the frequency of vowel reduction, or the frequency of coda deletion) based on contributions from a number of other *explanatory factors* (Agresti, 1996). Thus when we report that an effect was significant, it is meant to be understood that it is a significant parameter in a model that also includes the other significant variables. In other words, after accounting for the effects of the other explanatory variables, adding the explanatory variable in question produced a significantly better account of the variation in the response variable. For duration, which is a continuous variable, we used ordinary linear regression to model the log duration of the word. For vowel quality and coda deletion, which are categorical variables, we used logistic regression.

All of the analyses assume that the items of the sample are independent. This is not strictly true, since some of the function words come from the same pseudo-utterances (and more generally, from the same conversations). We do not expect that this would influence the results much, but it does mean that the significance levels we report are somewhat overstated, and hence they should be interpreted with caution.

2.3 Control factors

The reduction processes are each influenced by multiple structural and performance factors that must be controlled to assess the contribution of lexical category to reduction. We briefly review these factors here and our method of controlling for them. First, we excluded tokens of *of*, *you*, *to*, and *that* based on the following three factors:

special forms: We excluded cliticized words (e.g., *you've*, *that's*, etc.), Such forms made up about nine percent of the total occurrences of the four function words; tokens of *that's* accounted for almost 80 percent of the excluded items. Also excluded were about 100 items whose extreme or aberrant parameter values indicated a high likelihood of coding error.

prosodic position: We removed words which began or ended the pseudo-utterances of our dataset, because of the incomparability of the predictability measures for such items and their high correlation with special prosodic contexts. Such words made up about 14 percent of the items. Recall that these conversational fragments were bounded either by turns or long pauses. Thus all turn-initial and turn-final items were excluded. Many items that were initial or final in an intonational phrase would also have been excluded, leaving only such items that fell within the pseudo-utterances. Based on two subsamples of the data coded for prosodic units, we estimate that perhaps 10 percent of the remaining data consists of items that were initial or final in the intonational phrase.

planning problems: Previous work has shown that when words are followed or preceded by disfluencies indicating planning problems (pauses, filled pauses *uh* and *um*, or repetitions), their pronunciations are less reduced (Fox Tree & Clark, 1997; Jurafsky et al., 1998; Bell et al., 1999; Shriberg, 1999). Partly for this reason and partly because the interpretation of the predictability variables in such contexts was unclear, these items, about 18 percent of the remaining data, were excluded.

We then controlled other variables known or suspected to affect reduction by entering them first in the regression model. Thus the full base model for an analysis was a regression on the following set of control factors:

rate of speech: Speech researchers have long noted the association between faster speech, informal styles, and more reduced forms. For a recent quantitative account of rate effects in Switchboard, see Fosler-Lussier and Morgan (2000). We measured rate of speech at a given function word by taking the number of syllables per second in the smallest pause-bounded region containing the word. Our regression models all included log rate; log squared rate, found to be a significant factor in our work with larger samples, was included in models where it was an appreciable factor.

segmental context: A general fact about reduction processes is that the form of a word is influenced by the segmental context—for example, consonant deletion is favored when a segment is preceded by or followed by a consonant. We controlled for the class (consonant or vowel) of the following segment.

For vowel reduction, we also controlled for whether the target syllable was open or closed (e.g., *it* vs. *to*), since we know from studies in larger samples that this variable interacts closely with segmental context (the later factor was ignored in some regressions where its effect was negligible).

Reduction of following vowel: The prosodic pattern of the utterance plays a crucial role in reduction. Since our current dataset does not mark stress or accent, the only prosodic control was whether the vowel in the syllable following the target word was reduced or full. (This partially controls for stress since the reduction of the following vowel should correlate with its stress level, and hence the stress level of the target word.)

Probability of word given neighboring words: In earlier work (Jurafsky et al., 2001), we showed that the conditional probability of a word given the previous and following words played an important role in its reduction. We therefore included four probabilistic control factors: the conditional probability of the target word given the previous word, the joint probability of the target word with the previous word, the conditional probability of the target word given the following word, and the joint probability of the target word given the following word. The next section summarizes the definitions of these probabilities and the effects we found in this earlier work.

We also included terms for some of the interactions between these variables where their effect was appreciable.

Several factors that have been reported to influence reduction were not controlled in this study. First, our definition of words was quite simplified; we assume that anything bounded by spaces in the text transcriptions was a word. Thus *most of*, for example, was taken to be two words.

Other factors not controlled included the segmental environment of the preceding word, finer details of the the segmental environment of the following word, i.e. register, age, sex, and social class. We expect that the remaining segmental factors would have relatively little effect on duration or on the role of the predictability measures, but we have not examined this possibility. When we controlled for some social variables in earlier work (Bell et al., 1999), the effects on reduction were relatively small and the robust effects of the predictability measures were little diminished. The effects of prosodic structure, stress, and accent are only partially and indirectly controlled by the variable of reduction in the following vowel and the exclusion of beginnings and ends of pseudo-utterances. Work in progress in which we controlled for an approximation to position in the intonational phrase showed that while initial and final tokens in this domain were significantly longer, controlling for these items had little effect on the predictability measures. The effect of pitch accent is discussed below in section 3.4.

2.4 Effects of probability of word given neighboring words

Jurafsky et al. (2001) proposed the *Probabilistic Reduction Hypothesis*: Word forms are reduced when they have a higher probability. In that paper we showed specifically that target words which have a higher probability given neighboring words are shorter. We use two measures (the *joint probability* and the *conditional probability*) of the predictability of a word given the previous and given the following word. The *joint probability* of two words $P(w_{i-1}w_i)$ may be thought of as the prior probability of the two words taken together, and is estimated by just looking at the relative frequency of the two words together in a corpus:

$$P(w_{i-1}w_i) = \frac{C(w_{i-1}w_i)}{N} \tag{1}$$

The *conditional probability of a word given the previous word* is also sometimes called the *transitional probability* (Bush, 1999; Saffran et al., 1996). The conditional probability of a particular target word w_i given a previous word w_{i-1} is estimated from a sufficiently large corpus, by counting the number of times the two words occur together $C(w_{i-1}w_i)$, and dividing by $C(w_{i-1})$, the number of times that the first word occurs:

$$P(w_i|w_{i-1}) = \frac{C(w_{i-1}w_i)}{C(w_{i-1})} \tag{2}$$

The difference between the conditional and joint probability is that the conditional probability controls for the frequency of the conditioning word. For example, pairs of words can have a high joint probability merely because the individual words are of high frequency (e.g., *of the*). The conditional probability would be high only if the second word was particularly likely to follow the first. Most measures of word cohesion, such as conditional probability and mutual information, are based on such metrics which control for the frequencies of one or both of the words (Manning & Schütze, 1999).

In addition to considering the preceding word, we measured the effect of the following word by the two corresponding probabilities. The *joint probability of a word with the next word* $P(w_iw_{i+1})$ is estimated by the relative frequency of the two words together:

$$P(w_iw_{i+1}) = \frac{C(w_iw_{i+1})}{N} \tag{3}$$

Similarly, the *conditional probability of the target word given the next word* $P(w_i|w_{i+1})$ is the probability of the target word w_i given the next word w_{i+1}. This may be viewed as the predictability of a word given the word the speaker is about to say, and is estimated as follows:

$$P(w_i|w_{i+1}) = \frac{C(w_i w_{i+1})}{C(w_{i+1})} \tag{4}$$

Jurafsky et al. (2001) showed that all of these measures played a role in reduction; words which were highly probable by any of these measures were shorter, more likely to have a reduced vowel, and more likely to have a deleted consonant.

Since our 38,000 word corpus was far too small to estimate word probabilities, we used the entire 2.4 million word Switchboard corpus (from which our corpus was drawn) instead. See Jurafsky et al. (1998) for details about the backoff and discounting methods that we used to smooth the estimates of very low frequency items. We then took the log of these probabilities for use in our regression analyses.

2.5 Earlier results on sensitivity of our methodology to lexical frequency

The hypothesis that more frequent forms are more likely to be reduced in lexical production has been widely proposed (Schuchardt, 1885; Jespersen, 1922; Zipf, 1929; Martinet, 1960; Oldfield & Wingfield, 1965; Fidelholz, 1975; Hooper, 1976; Phillips, 1984; Jescheniak & Levelt, 1994; Bybee, 2000; Rhodes, 1992, 1996).

In Jurafsky et al. (2001), we used our corpus-based methodology to examine the role of various probabilistic measures, including lexical frequency, on reduction. We examined 2042 tokens of words ending in t or d, again from the 38,000 word phonetically-transcribed Switchboard database. We examined two dependent measures: the duration of the word in milliseconds (only for monosyllabic words) and deletion of the final t or d. We found a strong effect of relative frequency on both kinds of reduction ($p < .0001$). For duration, high frequency words (at the 95th percentile of frequency) were 18% shorter than low frequency words (at the 5th percentile) and high frequency words (at the 95th percentile) were 2.0 times more likely to have deleted final t or d than the lowest frequency words (at the 5th percentile).

These results suggest that our corpus-based methodology is indeed sensitive to lexical frequency. But our previous results did not distinguish between lemma frequency and wordform frequency. Testing the differential role of the two requires labeling each wordform observation for its associated lemma. This coding process is described in the next section.

2.6 Coding for lexical category

We coded the four function words *that*, *to*, *of*, and *you* for their syntactic categories. In each case we coded fine-grained categories which were then collapsed into broader categories for our analyses. Tables 3–5 show examples of the different lexical categories we coded for the four words.

Count	%	Syntactic Category	Example from Switchboard
543	74%	**infinitive marker**	is that a tough system **to** be in?
195	26%	**preposition/particle**	she is a great comfort **to** me
738	100%	**Total**	

Table 2: Lexical category coding for *to*.

For *that*, there are four traditional part-of-speech or syntactic category differences: determiner, pronoun, complementizer, and relativizer. Some previous studies of form variation have treated complementizers and relativizers together (Jespersen, 1933), while others have distinguished all four (Berkenfield, 2000). We chose to look at all four. We did not, however, study subtypes of these categories such as the differences between subject and object pronouns, or between subject and object relatives, mainly since neither the literature nor our intuitions gave us any reason to expect a duration or reduction difference between subject and object pronouns. But in order to make it clear what the categories encompass, we give more details of subtypes of these categories in Table 3.

Count	%	Syntactic Category	Example from Switchboard
294	37%	**pronoun**	
		subject (105)	**that** didn't help at all
		non-subject (189)	and we keep thinking about **that**
183	23%	**complement**	
		verbal complement	like uh one company had proposed to me **that** i could come back to work after having the baby um
		verbal extraposition	and uh it's always occurred to me **that**
		nominal complement	I just finished fuming at the fact **that** we pay an eight and a half percent sales tax ...
170	21%	**rel. pronoun**	
		subject relative (101)	found a bunch of memos **that** were uh supposedly from...
		non-subj. relative (69)	you get on a topic **that** you know you enjoy...
102	13%	**determiner**	...and fines and things of **that** nature
42	5%	**other**	idioms (*that is*), intensifiers, etc.
791	100%	**Total**	

Table 3: Lexical category coding for *that*.

For *of*, the different lemmas are strongly related to the different syntactic constructions they can occur in, for example as the complement of a verb or preposition (e.g., *thought* **of**, *outside* **of**) or as a partitive (e.g., *some* **of**, *all* **of**) (Table 4).

Count	%	Syntactic Category	Example from Switchboard
230	40%	**partitive**	e.g., one **of** them, type **of** job, all **of**, some **of**
100	18%	**complement**	e.g., thought **of**, outside **of**, in front **of**
95	17%	**genitive/other postnominal**	e.g., friend **of** mine, matter **of** concern, things **of** that nature
146	26%	**assorted idioms**	e.g., kind **of**, lot **of**, sort **of**, matter **of** fact
561	100%	**Total**	

Table 4: Lexical category coding for *of*.

For *you*, we distinguished four potential lemmas. We considered the *you* of the phrase *you know*, the referential pronoun *you*, and a non-referential or generic pronoun *you* (*yeah then* **you** *have to get up the next day and move it on*). We also distinguished auxiliary-inverted and non-inverted instances of the referential *you* (Table 5).

Count	%	Syntactic Category	Example from Switchboard
359	47%	**you know**	it was **you** know in the seventies
212	27%	**generic**	yeah then **you** have to get up the next day and move it on
172	22%	**referential**	
83	11%	aux-inverted	well do **you** drink soda and such?
89	12%	not aux-inverted	and **you** only get one of them
743	100%	**Total**	

Table 5: Lexical category coding for *you*.

3 Results

We ran separate regressions for each of the four words.

3.1 Effect of the lemma on variation in *to*

As Table 2 shows, the frequency of the infinitive lemma was 2.8 times greater than the frequency of the preposition lemma. If lemma frequency plays a role in lexical production, as the lemma frequency hypothesis suggests, we would expect the infinitives to be shorter than the prepositions after controlling for other variables in our regression. In addition, we would expect the infinitives to have a greater percentage of reduced vowels than the prepositions. If we look at the values of tokens of the two lemmas in the corpus after excluding items but not controlling for any other factors, this is indeed the case, as we see in Table 6.

Lemma	Count	Proportion of Occurrences (percent)	Average Duration (ms)	Percent Vowel Reduction
Infinitive	543	74	109	78
Preposition	195	27	138	56

Table 6: Lemma counts, proportions, raw durations, and raw percentage of reduced vowels for infinitival and prepositional *to*. Lemma counts and proportions are based on the total sample. Durations and reduced vowel percentages are based on the sample after excluding non-comparable items.

The differences in all three variables are highly significant ($p < .0001$), and are in the direction predicted by the lemma frequency hypothesis. We used a binomial test for the lemma frequency difference ($z_c = 7.2$). To examine the differences in duration and in vowel reduction of the two categories, we made a planned comparison between the infinitive and the preposition categories in linear regressions for duration, and in logistic regressions for vowel reduction. These comparisons yielded values of $F(1, 567) = 28.4$ for duration, and $\chi^2(1) = 21.9$ for vowel reduction.

These differences, however, are not controlled for factors known to influence reduction. We thus added control variables for rate of speech and for following segmental context to the regression models. There remained a difference in duration ($F(1, 564) = 12.1, p = .0005$), and in vowel reduction ($\chi^2(1) = 5.7, p < .05$) between infinitival and prepositional *to*. The infinitives were shorter in duration and had a greater probability of having a reduced vowel.

Still not controlled in these comparisons are the effects of local predictability. We therefore included four local predictability variables into our base model for the regression:

1. Joint probability of target and previous word
2. Joint probability of target and following word
3. Conditional probability of target given following word
4. Conditional probability of target given previous word

After adding these variables, there remained no effect of the lemma categories on either duration or vowel reduction ($F < 1, \chi^2 = 1.1$). That is, there was no difference between the duration or vowel quality of infinitive *to* and preposition *to*. This was a surprising result, as the difference in raw durations was so large.

3.2 The roles of predictability and lemma category

The fact that no effect of lexical category remains after controlling for predictability implies that there must be some connection between lemma category and predictability. And indeed there is. The average conditional probability of prepositional *to* given the following word is 0.032, whereas the average conditional probability of infinitival *to* is 0.190. That is, infinitival *to* is more predictable. We might further inquire, however, about the symmetry of the relationship. Are lexical category and predictability here merely correlated, so that either is an equally good predictor of reduction?

Or is predictability a separate factor, whether we categorize the data by lemma or not? This latter model is what was suggested by our earlier work in which we found a robust effect of predictability on reduction throughout a very wide range of word classes, word frequencies, and word contexts.

We tested whether predictability is a separate factor by examining the effect of predictability after controlling for lemma. As we expected, the predictability variables remain highly significant. This means that predictability completely accounts for any predictive power that the lemma variable offers, and offers further explanatory power, e.g. it accounts for the reduction of tokens within each of the two lemma categories. This suggests that predictability and not lemma is the factor accounting for the difference between surface pronunciations of *to*.

In other words, the lemma categories only appeared to account for the reduction and duration difference for *to* because it happens that the more reduced infinitival *to*s are also more predictable.

3.3 Lexical categories of *of*

Recall that the lexical categories for *of* included complements (*thought of*), partitives (*one of them*), genitives/postnominals (*friend of mine*), and high-frequency idioms *kind of*, *sort of*). Their frequencies, average durations, and average coda deletion are summarized in Table 7. The lemmas did not differ in frequency of vowel reduction.[2]

The frequency of partitives is obviously much greater than that of complements and genitive/postnominals, whose frequencies do not differ significantly.

To examine the differences in the durations and coda deletion of the categories, we made planned comparisons between the frequent partitive and the other two categories, and then between the latter two categories, using contrast variables in linear regressions for duration, and in logistic regressions for coda deletion. Partitives

[2] The two frequent idioms *kind of* and *sort of*, which constituted 12% of the tokens, are not compared here, since they may well be acting as a lexicalized category on their own, rather than playing the part of partitives or genitives. However, the duration of *of* in these idioms is shorter (74 ms) and they have more frequent deletion of the coda (64%).

Lemma	Count	Proportion of Occurrences (percent)	Average Duration (ms)	Percent Coda Deletion
Partitive	230	54	82	55
Complement	100	24	94	39
Genitive/Postnominal	95	22	103	38

Table 7: Lemma counts, proportions, raw durations, and raw percentage of deleted codas for *of*. Lemma counts and proportions are based on the total sample. Durations and reduced vowel percentages are based on the sample after excluding non-comparable items.

differ from complements/genitives both for raw duration ($F(1,474 = 16.6, p = .0001)$) and raw proportion of coda deletion ($\chi^2(1) = 13.7, p < .0005$). The differences between complements and genitives are not significant. Note that the lemma frequency hypothesis predicts that partitives should be shorter and more reduced than complements and shorter and more reduced than genitives, which is the case.

When we control for rate and contextual factors (without predictability), both duration ($F(1,471 = 14.9, p = .0001)$) and coda deletion ($\chi^2 = 11.2, p < .001$) remain significant. Finally, controlling also for predictability, we still found effects of lemma on coda deletion, but not on duration ($F < 1$). Partitives were marginally more likely to have deleted codas than the combination of complements and genitives ($\chi^2(1) = 5.1, p < .05$).[3]

These results suggest that surface pronunciations of the (more frequent) partitive lemma for *of* are more reduced than surface pronunciations of the (less frequent) complement or genitive lemmas. While the predictability variables completely eliminated any lemma effect for *to*, this was not the case for *of*.

3.4 Lexical categories of *that*

The frequencies, average durations, average reductions, and average coda deletion of the four major lexical categories of *that* are summarized in Table 8.

From Table 8 we observe that the pronoun *that* is the most frequent of the four lemmas of *that* that we consider, and the determiner or demonstrative *that* is the least frequent, with relative marker *that* and complementizer *that* falling in between. All the frequency differences between the categories are significant except for that between complements and relative markers (z_c from 4.1 to 9.6, $p < .0001$). Berkenfield (2000) investigated the durations of these same four lemmas for 305

[3] In all of these comparisons, we included the frequent and probably lexicalized *lot of* in the partitive category. These tokens made up 20% of the total sample of partitives. Analyses on this smaller restricted set of partitives showed little difference from those reported here.

Lemma	Count	Proportion of Occurrences (percent)	Average Duration (ms)	Percent Vowel Reduction	Percent Coda Deletion
Pronoun	294	37	186	2	56
Relative	170	21	132	18	50
Complementizer	183	23	154	22	43
Determiner	102	13	142	0	75

Table 8: Lemma counts, proportions, raw durations, and raw percentage of reduced vowels and deleted codas for *that*. Lemma counts and proportions are based on the total sample. Durations and deleted coda and reduced vowel percentages are based on the sample after excluding non-comparable items.

observations of *that* in a corpus of conversational speech taken from the television program *The Newshour with Jim Lehrer*. She found a similar ranking of the raw durations, although not quite the same. In her data, as in ours, pronouns were longest. But in her data determiners were the next longest, followed by complementizers, with relative clause markers the shortest. That is, in her data, determiners and relative clause markers are switched in their order from our Switchboard data. Berkenfield (2000) also found that while the pronoun lemma for *that* was the most frequent in Switchboard, it was the complementizer lemma that was by far the most frequent in written corpora like the Brown corpus (Francis & Kučera, 1982), and news programs like the Lehrer Newshour. Thus it is not clear what the correct prediction would be of the lemma frequency hypothesis.

All corpora agree, however, that the determiner lemma is the least frequent. It is also less likely to have a reduced vowel than complementizers or relative markers. This difference cannot be attributed to rate, contextual factors, or predictability; it remains highly significant after controlling for these factors ($\chi^2(1) = 23.5, p < .0001$). The duration and likelihood of coda deletion of determiners, however, do not differ significantly from complementizers or relative markers.

Pronouns appear to be much like determiners in that they, too, are less likely to have reduced vowels than the complementizers and relative markers. But unlike determiners, they are longer than the other lemmas (including determiners) after accounting for the control factors ($F(1, 421) = 20.7, p < .0001$). They do not differ in likelihood of coda deletion.

Complementizers are less reduced than relative markers: they are longer ($F(1, 419) = 9.2, p < .005$) and marginally less likely to have deleted codas ($\chi^2(1) = 5.1, p < .05$); but they do not differ in likelihood of vowel reduction.

It may be that the occurrence of accent is an important factor in these results. While we do not have enough data coded for accent to control for this factor, a small portion of the Switchboard corpus has been coded for accent under the direction of Stefanie Shattuck-Hufnagel and Mari Ostendorf. They generously made an alpha-release of their accent-coded corpus available to us, and we examined the overlap between their corpus and ours, a small subset consisting of 10 percent or less of our entire dataset. There were 180 tokens of our four function words in this sample, if we include the disfluent contexts. Of these, 16 were accented. Only *that*s received accent at all frequently, 10 out of the 36 tokens coded; *Of*s were accented once out of 45 tokens; *to*s, once out of 53; *you*s, 4 times out of 46. This is encouraging in that it affords some confidence that the results for *of* and *to* are unlikely to be much influenced by accent.

Of the 36 tokens of *that* coded for pitch accent, only the determiners (4 accented of 11) and the pronouns (6 of 14) received accent; not surprisingly, all accent-coded occurrences of complementizers and relative markers were unaccented. However, it appeared that the disfluent contexts favored accent somewhat. In fluent contexts, only two of eight determiners and two of nine pronouns were accented. Although, for example, the average duration of the accented pronominal *that*s was longer than the duration of the unaccented ones (256 ms versus 142 ms), the sample is simply too small to demonstrate significant differences of duration either for the accent or lemma categories.

Until the influence of accent can be further determined, any conclusions about pronunciation differences linked to *that* lemmas must remain guarded. The lesser reduction found for determiners is surely affected to some extent by the presence of accented items, and is in any case suspicious without any accompanying difference in duration. At least for pronouns there are strong effects both for reduction and duration, but until the factor of accent can be controlled, we can only conjecture that a noncontextual effect may exist. The relatively weak difference in duration between complementizers and relative markers remains, however, since accent is unlikely to account for any difference between them.

3.5 Lexical categories of *you*

Unlike the other three words, the two main lemmas we investigated for *you* (the referential and the generic/nonreferential) did not differ significantly in raw duration or reduction. None of the differences in duration or vowel reduction between generic and referential lemmas shown in Table 9 are significant.

After controlling for the base model including predictability, the items were still the same; referential and non-referential observations of *you* did not differ in surface pronunciation.

Lemma	Count	Proportion of Occurrences (percent)	Average Duration (ms)	Percent Vowel Reduction
You know	359	47	105	49
Generic, non-referential	211	27	109	26
Referential	172	22	118	22

Table 9: Lemma counts, proportions, raw durations, and raw percentage of reduced vowels for *you*. Lemma counts and proportions are based on the total sample. Durations and reduced vowel percentages are based on the sample after excluding non-comparable items.

4 Summary

For the word *to,* we found no separate effect whatsoever of the lemma for word-form variation. That is, we could account for most or all of the variation in surface form of *to,* solely based on the control factors described above. The most important such control factor was the predictability of the word given the neighboring words. When this factor was included in our regression, it accounted for all of the differences between the different surface realizations of *to*.

While we found no evidence for a lemma effect for *to*, we did find a lemma effect for *of*, even after controlling for predictability. The partitive *of* (*one* **of** *them*) is more likely to have a deleted coda than the genitive *of* (*friend* **of** *mine*) and the complement *of* (*to think* **of**). Since the partitive is also more frequent, the direction of the difference is consistent with the hypothesis that the more frequent lemma will show more reduction.

Our results on *that* are inconclusive. First, after controlling for predictability and other factors, the pronoun *that* (*thinking about* **that**) is longer and more likely to have a full vowel than the relative (*a topic* **that** *you enjoy*) or complementizer *that* (*proposed to me* **that** *I could*). Second, the determiner *that* (*things of* **that** *nature*) was more likely to have a full vowel than the complementizer or relative *that*. Both these differences, though, are clearly related to prosody; the pronoun *that* and the determiner *that* are both much more likely to receive accent than the complementizer or relative marker.

The third result with *that* is less likely to be influenced by accent. The relativizer sense of *that* is shorter and has more coda deletion than the slightly less frequent complementizer sense of *that*. This difference is unlikely to be caused by lemma frequency, since the relative frequencies of the complementizer and relativizer lemmas were not significantly different.

Finally, we found no effects of lemma on the word *you*.

5 Conclusions

We investigated the role that lemmas play in wordform variation, using a corpus-based methodology which is sensitive to lexical frequency effects. Raw duration and reduction measures seem to show differences between the surface forms corresponding to different lemmas for the words *of*, *that*, and *to*. But after controlling for such factors as rate of speech, segmental context, neighboring disfluencies, and, crucially, predictability from neighboring words, almost all of these differences disappeared. For example, the difference between infinitive and prepositional lemmas of *to* turned out to be explained by word-predictability factors that we had shown to predict reduction in earlier work (Jurafsky et al., 2001). In summary, after appropriate controls, we found no differences whatsoever in surface form for *you* or *to* that could be attributed to lemmas. There were also no differences due to lemmas in duration or vowel reduction for *of*.

Although we were able to account for most of the differences between surface forms of different lemmas via these contextual factors, four differences remained significant even after controlling for these factors. As the previous section suggested, two of these, the less reduced forms of the determiner and pronominal lemmas for *that*, may turn out to actually be effects of pitch accent.

Thus two differences remained. First, the frequent partitive sense of *of* was more likely to have a deleted coda than the less frequent genitive or complement senses. Second, the relativizer sense of *that* is shorter and has more coda deletion than the equally frequent complementizer sense of *that*.

The fact that these two differences remained despite our control for context suggests that the contextual model of lexeme variation may not be sufficient to explain pronunciation variation. The contextual model predicts that all variation in surface pronunciation should be accounted for by context or by production factors, and that neither the lemma nor the lexeme should have a role. This model is not sufficient to explain the variation we see in *of* and *that*.

But the differences in *of* and *that* are also not compatible with the lemma frequency model of form variation. The relativizer and complement lemmas for *that* have the same frequency, but still differ in duration and coda deletion percentage. Even the result for *of*, where the more frequent partitive had more coda deletion than the rarer genitive and complementizers, is not strong evidence for the lemma frequency model. This is because if the reduction of *of* is caused by frequency, we would also expect reduction differences between other lemmas whose frequency difference is as great as the difference in *of*. The frequency ratio between the partitive and complement/genitive lemmas for *of* is 2.4 to 1. But the frequency ratio between the infinitive and preposition lemmas for *to* is about the same (2.8 to 1), but *to* shows no lemma-based effect on reduction.

Even if in time numerous instances of association of frequent lemmas with reduced non-contextual pronunciations were found, it does not follow that production models should incorporate the lemma frequency hypothesis. The association would more plausibly be accounted for by diachronic preferences rather than synchronic structures. A homonym split could begin with a differentiation of pronunciations that was at first purely contextual, with reduced forms occurring in more frequent, more predictable, and possibly less prosodically prominent constructions or contexts. If the contextual differences became lexicalized, this would lead to some of the distinctions in reduction becoming encoded in the lexicon, and would leave an association of more reduction with the more frequent lemmas.

Our conclusion that the lemma frequency hypothesis does not hold for speech production is nicely consonant with the result of Jescheniak and Levelt (1994). But it does point out an intriguing difference between lexical access in comprehension, where lemma frequency effects are robust and have been reported cross-linguistically (Li & Yip, 1996; Ahrens, 1998), and lexical access in production (with, it seems, no lemma frequency effect). Lemma frequency, then, is a feature that seems to play different roles in language comprehension and production, a fact that should clearly be further studied.

Even if there is not sufficient support for the lemma frequency hypothesis, it is still necessary to consider whether there are cases which require the lemma-based model involving other factors than frequency. Diachronically, we assume such differences would arise by lexicalization of contextual differences. The issue is whether the differences are incorporated in the lexeme or in the lemma, i.e. whether the multiple lexeme model or the lemma-based model applies.

Recall that the multiple lexeme model suggested that different lemmas are differentially linked in the lexicon to different wordforms. Perhaps, for example, as shown in Figure 4, all lemmas for *of* are linked to the wordforms [ʌv] and [ʌ], but the genitive or complement *of* is linked more preferentially to the wordform [ʌv], while the partitive *of* is linked more preferentially to [ʌ]. Similarly, the relativizer sense of *that* could be preferentially linked to [ðæ], while the complementizer sense of *that* could be preferentially linked to [ðæt].

The multiple lexeme model might be able to handle such results. But recall that the relativizer sense of *that* is also *shorter* than the complementizer sense of *that*, even after controlling for the greater incidence of coda reduction. That is, the difference between these two forms is not just purely representable as a segmental difference. This suggests that the specifications of wordforms includes more phonetic detail than just phonological categories. We found a related result in our earlier work (Jurafsky et al., 2001), where we showed that reduced forms of very predictable words are shorter even after controlling for segmental changes (vowel reduction or coda deletion).

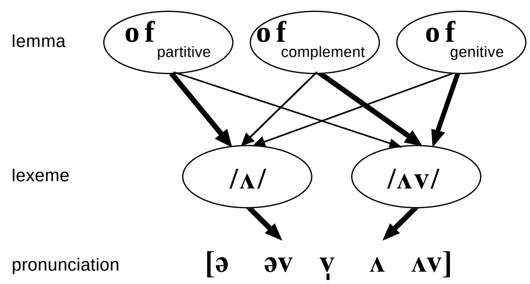

Figure 4: A possible multiple lexeme model of *of*, showing more coda deletion in partitive *of*, and less coda deletion in genitive and complement *of*.

One recent approach to phonological representation and production does offer a possible solution to these data. Pierrehumbert (2001) and others have recently proposed exemplar-based models of phone and word production. In exemplar-based models, each category is stored in memory as a cloud of exemplars of that category. Thus in Pierrehumbert's model, for example, phones and words are both stored as clouds of exemplars. Production of a phone or word takes place by randomly activating and then producing a particular exemplar. Pierrehumbert (2001) then proposes that the production process is not quite random; in leniting contexts it has a very slight bias in its selection process toward shorter forms. This models the general historic tendency of words to lenite. Pierrehumbert's model also predicts the effects of wordform frequency on reduction. A more frequent wordform will produce more exemplars, each of which is very slightly biased toward reduction. Over time, the exemplar cloud of a very frequent word will tend to consist of somewhat more reduced exemplars.

Such an exemplar-based model can explain the non-segmental nature of the differences we show in the production of *that*. Two lemmas, say the complementizer and relativizer *that*, may begin with a direct mapping to a single wordform. Over time, the lemmas may begin to be differentially mapped to different wordforms. These wordforms may have a clear segmental difference, as shown in Figure 2 and Figure 4, or the two wordforms may be segmentally identical. Either way, these different wordforms would themselves consist of clouds of exemplars. The exemplars for the more frequent wordform would in general be more reduced than the exemplars for the other; some of these reductions might be segmental, but many might simply consist of slightly shorter durations for the individual phones in the word. Such an exemplar-based model might also model our word predictability effects by including some exemplar clouds for two-word or three-word phrases.

While this exemplar-based explanation is clearly only a vague and preliminary attempt at a model, it is an exciting possibility which we are studying further.

Such examples suggest that it will always be possible to explain any noncontextual variation with a multiple lexeme model, provided that it is sufficiently elaborated. If so, there would be no need to resort to lemma-based models. The result of our study of lemma variation, therefore, is that there are cases which appear to require more complex representations at the lexeme level than have been commonly assumed. Ultimately, of course, we require independent evidence, from controlled experiments or other sources, for models like that sketched in Figure 4, or for models to account for variation at a finer level of detail than traditional phonological categories.

We are currently working on adding further prosodically-coded data to the accent-coded portion of Switchboard coded by Stefanie Shattuck-Hufnagel and Mari Ostendorf, so as to be able to reanalyze the pronoun and determiner senses of *that* after controlling for accent.

In addition to these conclusions about the process of lexical production, we would like to end with a methodological insight. We hope to have shown that a corpus-based methodology such as ours can augment traditional controlled psycholinguistic experiments to help provide insight into psychological processes like lexical production. Corpus-based methods have the advantage of ecological validity. The difficulty with corpus-based methods, of course, is that every possible confounding factor must be explicitly controlled in the statistical models. This requires time-consuming coding of data and extensive computational manipulations to make the data usable. Creating a very large hand-coded corpus is difficult, as we saw with our inability to completely control for pitch accent for the word *that*. But when such control is possible, a corpus provides natural data whose frequencies and properties may be much closer to the natural task of language production than experimental materials can be. Obviously, it is important not to rely on any single method in studying human language; corpus-based study of lexical production is merely one tool in the psycholinguistic arsenal, but one whose time, we feel, has come.

Acknowledgements

This project was partially supported by the National Science Foundation via NSF IIS-9733067 and IIS-9978025. Many thanks to Michelle Gregory, William D. Raymond, Eric Fosler-Lussier, Joan Bybee, and Janet Pierrehumbert for fruitful discussions. We are also grateful to Stefanie Shattuck-Hufnagel and Mari Ostendorf, who generously took the time and effort to release to us a preliminary version of their prosodically coded portion of Switchboard. Finally, we owe a particular debt of gratitude to one anonymous reviewer and to the editors of this volume, all of whom gave extraordinarily helpful comments and spotted many errors and inconsistencies which greatly improved our paper. Of course all remaining errors are our own.

References

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. John Wiley & Sons, New York.

Ahrens, K. V. (1998). Lexical ambiguity resolution: Languages, tasks, and timing. In Hillert, D. (Ed.), *Syntax and Semantics, Volume 31: Sentence Processing: A Crosslinguistic Perspective*. Academic Press, San Diego.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., & Gildea, D. (1999). Forms of English function words – Effects of disfluencies, turn position, age and sex, and predictability. In *Proceedings of ICPhS-99*, pp. 395–398.

Berkenfield, C. (2000). The role of syntactic constructions and frequency in the realization of English *that*. Master's thesis, University of New Mexico, Albuquerque, NM.

Bush, N. (1999). The predictive value of transitional probability for word-boundary palatalization in English. Master's thesis, University of New Mexico, Albuquerque, NM.

Bybee, J. L. (2000). The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M., & Kemmer, S. (Eds.), *Usage-based Models of Language*, pp. 65–85. CSLI, Stanford.

Ferguson, C. A. (1975). "Short a" in Philadelphia English. In Smith, E. (Ed.), *Studies in Linguistics in Honor of George L. Trager*, pp. 259–274. Mouton, The Hague.

Fidelholz, J. (1975). Word frequency and vowel reduction in English. In *CLS-75*, pp. 200–213. University of Chicago.

Fosler-Lussier, E., & Morgan, N. (2000). Effects of speaking rate and word frequency on conversational pronunciations. To appear, Speech Communication.

Fox Tree, J. E., & Clark, H. H. (1997). Pronouncing "the" as "thee" to signal problems in speaking. *Cognition*, *62*, 151–167.

Francis, W. N., & Kučera, H. (1982). *Frequency Analysis of English Usage*. Houghton Mifflin, Boston.

Godfrey, J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *IEEE ICASSP-92*, San Francisco, pp. 517–520. IEEE.

Greenberg, S., Ellis, D., & Hollenback, J. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. In *ICSLP-96*, Philadelphia, PA, pp. S24–27.

Hogaboam, T. W., & Perfetti, C. A. (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, *14*, 265–274.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. (Ed.), *Current Progress in Historical Linguistics*, pp. 96–105. North Holland, Amsterdam.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 824–843.

Jespersen, O. (1922). *Language*. Henry Holt, New York.

Jespersen, O. (1933). *Essentials of English Grammar*.

Jones, D. (1947). *An English Pronouncing Dictionary*. E. P. Dutton and Company, New York.

Jurafsky, D., Bell, A., Fosler-Lussier, E., Girand, C., & Raymond, W. D. (1998). Reduction of English function words in Switchboard. In *ICSLP-98*, Sydney, Vol. 7, pp. 3111–3114.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J., & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure*. Benjamins, Amsterdam. To appear.

Labov, W. (1989). The exact description of the speech community: Short **a** in Philadelphia. In Fasold, R., & Schiffrin, D. (Eds.), *Language Change and Variation*, pp. 1–57. Georgetown University Press, Washington, D.C.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Science*, *22*(1), 1–75.

Levelt, W. J. M. (1983). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.

Li, P., & Yip, M. C. (1996). Lexical ambiguity and context effects in spoken word recognition: Evidence from Chinese. In *COGSCI-96*, pp. 228–232.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

Martinet, A. (Ed.). (1960). *Elements of General Linguistics*. University of Chicago Press, Chicago.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, *17*, 273–281.

Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320–342.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J., & Hopper, P. (Eds.), *Frequency and the emergence of linguistic structure*. Benjamins, Amsterdam. To appear.

Rhodes, R. A. (1992). Flapping in American English. In Dressler, W. U., Prinzhorn, M., & Rennison, J. (Eds.), *Proceedings of the 7th International Phonology Meeting*, pp. 217–232. Rosenberg and Sellier, Turin.

Rhodes, R. A. (1996). English reduced vowels and the nature of natural processes. In Hurch, B., & Rhodes, R. A. (Eds.), *Natural Phonology: The State of the Art*, pp. 239–259. Mouton de Gruyter, The Hague.

Roach, P. (1983). *English Phonetics and Phonology*. Cambridge University Press, Cambridge.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical cues in language acquisition: Word segmentation by infants. In *COGSCI-96*, pp. 376–380.

Schuchardt, H. (1885). *Über die Lautgesetze: Gegen die Junggrammatiker*. Robert Oppenheim, Berlin. Excerpted with English translation in Theo Vennemann and Terence H. Wilbur, (Eds.), *Schuchardt, the Neogrammarians, and the Transformational Theory of Phonological Change*, Athenaum Verlag, Frankfurt, 1972.

Shriberg, E. (1999). Phonetic consequences of speech disfluency. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS-99)*, San Francisco, Vol. I, pp. 619–622.

Simpson, G. B. (1981). Meaning dominance and semantic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, *20*, 120–136.

Simpson, G. B., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(1), 28–39.

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, *15*, 1–95.