UC Berkeley
UC Berkeley Previously Published Works

Title

Time and Thyme Are not Homophones: The Effect of Lemma Frequency on Word Durations in Spontaneous Speech

Permalink

https://escholarship.org/uc/item/12g3x6pr

Journal

Language, 84(3)

ISSN

0097-8507

Author

Gahl, Susanne

Publication Date

2008

DOI

10.1353/lan.0.0035

Copyright Information

Peer reviewed

TIME AND THYME ARE NOT HOMOPHONES: THE EFFECT OF LEMMA FREQUENCY ON WORD DURATIONS IN SPONTANEOUS SPEECH

SUSANNE GAHL

University of California, Berkeley

Frequent words tend to shorten. But do homophone pairs, such as *time* and *thyme*, shorten equally if one member of the pair is frequent? This study reports an analysis of roughly 90,000 tokens of homophones in the Switchboard corpus of American English telephone conversations, in which it was found that high-frequency words like *time* are significantly shorter than their low-frequency homophones like *thyme*. The effect of lemma frequency persisted when local speaking rate, predictability from neighboring words, position relative to pauses, syntactic category, and orthographic regularity were brought under statistical control. These findings have theoretical implications for the locus of frequency information in linguistic competence and in models of language production, and for the role of articulatory routinization in shortening.*

1. INTRODUCTION. Frequent words tend to shorten. But do words that sound alike, such as *time* and *thyme*, shorten equally if only one of them is frequent? This question is of interest for two lines of research. The first concerns models of language production. There is broadly shared agreement in the psycholinguistic literature that language production involves accessing two levels of lexical information. These two levels are commonly referred to as the LEMMA, which comprises a word's semantic and syntactic properties, and the PHONOLOGICAL FORM (Bock 1995, Dell 1986, Levelt 1989). Word frequency, according to a very influential model of language production (Levelt et al. 1999), is a property of a word's phonological form, not of its lemma. A corollary of that position is FREQUENCY INHERITANCE, the idea that homophone pairs, that is, pairs of words with identical form, should behave as though they had identical frequencies, namely the sum of the frequencies of all words sharing that form. If this is correct, and if word duration varies as a function of frequency, then homophones should shorten equally.

The second line of research concerns the place of frequency information in speakers' knowledge of language. Whether speakers' knowledge of usage frequency should be considered part of linguistic competence and whether usage frequency should even be an object of study for linguistics are issues that have given rise to disagreement. Interestingly, researchers on opposite sides of that disagreement have pointed to the link between word frequency and duration in support of their respective positions.

For example, Bybee (2001, 2002a,b, 2006) has argued that the shorter duration of frequent forms relative to infrequent ones is a way that frequency leaves its mark on the lexicon and thus constitutes evidence in favor of considering frequency information to be part of the lexicon. The mechanism that brings about shortening, according to Bybee, lies in articulatory routinization: 'With repetition, neuromotor routines become more compressed and more reduced' (2001:78). This appeal to articulatory routinization

has left usage-based and probabilistic approaches to linguistic competence open to criticism. For example, Newmeyer argues: 'It is a truism that the more often we do something, the faster we are able to do it. That is as true for language as for anything else' (2006:401). On that view, the shortening of frequent forms cannot be considered evidence that knowledge of usage frequency forms part of linguistic competence. If anything, the argument goes, the apparent similarity between shortening of frequent forms and motor-practice effects in other domains provides a reason AGAINST considering frequency information to be within the scope of linguistic knowledge.

Phonological pattern frequency indeed influences the speed of articulatory movements (Munson 2001). But is this increased motor fluency really the sole source of shortening? Homophone durations can shed light on this issue: if articulatory routinization is indeed the reason why frequent forms tend to be shorter than infrequent ones, then articulatory practice with a high-frequency word like *time* should cause a low-frequency homophone like *thyme* to shorten as well.

This article provides an analysis of the duration of roughly 90,000 tokens of homophones in the Switchboard corpus of American English telephone conversations (Deshmukh et al. 1998, Godfrey et al. 1992). To preview the results, high-frequency words like *time* were significantly shorter than their low-frequency homophones like *thyme*. Multiple linear regression models were used to examine the effect of lemma frequency while controlling for local speaking rate, predictability from neighboring words, position relative to pauses, syntactic category, and orthographic regularity. The effect of lemma frequency persisted when these other factors were controlled for. These results call into question both the phonological form as the sole locus of frequency information in the lexicon, and articulatory routinization as the prime source of frequency-dependent variation in duration. The results are consistent with a conception of language production in which frequency information pervades all levels of linguistic representations and mechanisms.

2. PREVIOUS LITERATURE.

2.1. FREQUENCY INHERITANCE. Language production is the result of many processes, including conceptual preparation, lemma selection, retrieval of word forms (or phonological forms), and articulation. On a frequency-based view of language, one might expect usage frequency to affect each of these component processes. However, in an influential paper, Jescheniak and Levelt (1994) advanced a much more constrained hypothesis, arguing that word frequency is encoded at the level of the phonological form, and not at the lemma level: '[w]ord form access is the major, and probably unique, locus of the word frequency effect' (Levelt et al. 1999:5).

Crucial evidence for that hypothesis comes from FREQUENCY INHERITANCE among homophones. The basis of frequency inheritance is that a given phonological form is activated every time any word associated with that form is activated. If the phonological form is the locus of frequency, then items sharing the same form should act as though they had identical frequencies, namely the sum of the frequency of all items sharing that form. For example, the frequency of the form /taɪm/ is the summed frequency of all associated items, in this case *time* and *thyme*: whenever either *time* or *thyme* is processed, their shared form is accessed. As a result, the low-frequency word *thyme* should in some respects behave like its high-frequency twin *time*.[1]

---

[1] A referee asks what effect having a low-frequency homophone would have on a high-frequency word. Since the form frequency is the sum of the frequencies of all words sharing that form, high-frequency words with low-frequency twins should behave as though they had slightly higher frequencies than their actual frequency. In practice, such an effect would most likely be undetectable, since, as pointed out in §6 below,

Two types of evidence lend support to the idea of frequency inheritance: production latencies, that is, the time it takes speakers to initiate speech, and speech errors. Jescheniak and Levelt explored the effects of frequency inheritance on production latencies in a translation task, in which participants were asked to produce the Dutch translations of visually presented English words (Jescheniak & Levelt 1994, experiment 6). There were three types of words: (i) low-frequency words with high-frequency homophones, (ii) low-frequency words without homophones (low-frequency controls), and (iii) high-frequency words without homophones (high-frequency controls). Jescheniak and Levelt found that production latencies of low-frequency words with high-frequency homophones were as fast as those of high-frequency words, consistent with the conclusion that the phonological form is the locus of frequency information in the lexicon.

Evidence for frequency inheritance originally came from patterns in speech errors (Dell 1990). Other things being equal, low-frequency words are more vulnerable to speech errors than high-frequency words. Dell hypothesized that low-frequency words with high-frequency twins might be less vulnerable to errors than low-frequency words without such twins. This prediction was confirmed.

It is important to realize that frequency inheritance is consistent with other mechanisms besides shared form frequency. In fact, frequency inheritance was originally hypothesized in the context of a model in which frequency information was located at the lemma level (Dell 1986, 1990). In that model, frequency inheritance arises because the highly activated phonological form of a high-frequency word sends activation to the lemma of a low-frequency homophone twin, as shown in Figure 1.
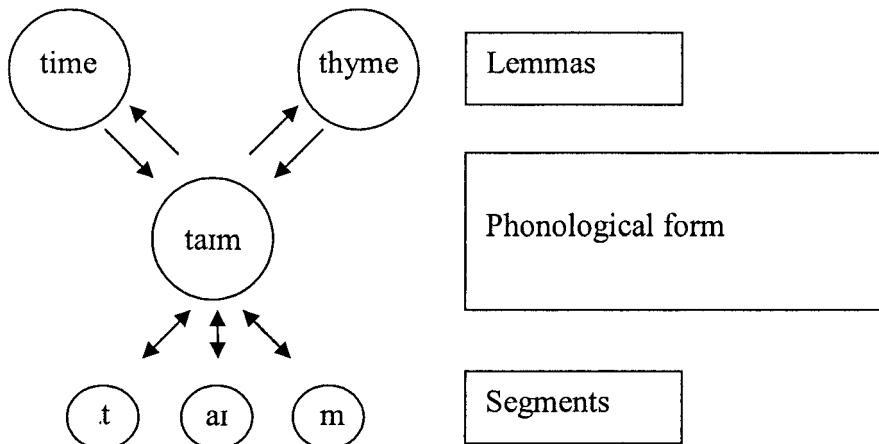

FIGURE 1. Schematic representation of activation flow between homophone pairs, loosely based on Dell 1990.

Frequency inheritance effects are not ubiquitous. Griffin (2002) points out that in interactive models of the type shown in Fig. 1, frequency inheritance might affect not just how quickly a word is accessed, but also whether the word is selected at all: speakers

---

the function relating word frequency to language processing appears to be approximately logarithmic, so that small absolute differences in frequency will have very minor effects, if any, in the high-frequency ranges. A further possibility raised by the referee is that high-frequency words might behave like low-frequency words by virtue of having low-frequency homophones. I am not aware of evidence supporting this idea.

should be more likely to produce low-frequency words if those words have high-frequency homophone twins than if they do not. Griffin's (2002) results did not bear out this prediction. However, Ferreira and Griffin (2003) found that priming of a homophone did affect subsequent lemma selection. Taken together, the results of Griffin 2002 and Ferreira & Griffin 2003 suggest that the existence of a homophone twin can affect the way a lexical item is processed, consistent with frequency inheritance, but that other factors may override such interactions between homophones.

Further evidence of frequency noninheritance comes from a series of studies by Caramazza and colleagues (2001). They conducted picture-naming experiments using English and Chinese stimuli in a design similar to that in Jescheniak & Levelt 1994, which yielded a failure to replicate Jescheniak and Levelt's findings (see Jescheniak et al. 2003 for discussion). On the basis of this and related findings, Caramazza and colleagues argue against postulating lemma representations as distinct from form representations (Caramazza 1997, Miozzo & Caramazza 2003, 2005). Under their alternative account, homophone pairs have fully independent representations, without a shared phonological form.

Another study reporting frequency noninheritance (Bonin & Fayol 2002) reported faster picture-naming latencies for high-frequency compared to low-frequency homophones in French. As Jescheniak and colleagues (2003:437) point out, it is possible that 'frequency inheritance is more likely to be observed in some languages than in others', perhaps due to differences in the lexical retrieval in languages with relatively many homophones, such as Chinese, compared to languages with relatively few homophones, such as Dutch.

In sum, homophones have provided evidence constraining models of lexical representation and word production. Some models lead one to expect lemma-frequency effects, whereas other models would need to be revised to accommodate such effects. An understanding of whether lemma frequency affects language production is important for all models of language production.

2.2. HOMOPHONES AND WORD DURATION. The evidence reviewed thus far comes from speech errors and production latencies, two important sources of evidence for the processes underlying language production. Word durations constitute a third potential source of evidence. Word durations reflect other factors besides those influencing latencies and speech errors, so it is conceivable that observations on word durations might diverge from findings in these other aspects of language production. But word durations, and the shortening of frequent words in particular, have proven a highly sensitive diagnostic for frequency effects in language processing (Bell et al. 2003, Bybee 2001, Fosler-Lussier & Morgan 1999, Jurafsky 2003, Jurafsky et al. 2001b, Krug 1998, Losiewicz 1995, Pluymaekers et al. 2005a,b). Therefore, if word durations reflect only effects of form frequency, never of lemma frequency, then this should cast doubt on the lemma level as a locus of frequency information.

It might be objected that homophones by definition sound alike and cannot provide a suitable diagnostic for effects of production processes on pronunciation. But pairs such as *laps* vs. *lapse* reportedly differ in duration as a function of their morphemic structure (Losiewicz 1995, Walsh & Parker 1983). There is also evidence of differences in the pronunciation of words with positive connotations, such as *bridal*, and homophonous words with neutral or negative connotations, such as *bridle* (Nygaard et al. 2002). Similarly, near-homophones may diverge historically as a function of word frequency, as argued by Phillips (2000) for pairs such as *tube/tuber* and *new/nude* in American

English. Therefore, supposed homophones can differ in pronunciation, in ways that may reflect processes underlying language production.[2]

Based on frequency inheritance, one would expect low-frequency words with high-frequency homophones to be as short as their high-frequency twins. Is this the case? The experimental record on this question is mixed. A majority of studies of homophone durations have so far failed to find differences in the durations of homophone pairs as a function of frequency. Other studies, however, found duration differences in some experiments, apparently varying with presentation order and context. Whalen (1991, 1996) found duration differences when homophones were presented in word lists with the words grouped by frequency, but not when the same words were presented in mixed-frequency lists. Similarly, Guion (1995) found that pairs of homophones differed in duration when the words were embedded in constructed sentence pairs (such as *We'll* **need** *the watch for a few hours*, *We'll* **knead** *the dough for five minutes*). When the same words were read in generic carrier phrases (*Say . . . to me again*), however, there was no significant difference in duration. Another study that did not report any significant durational differences between homophone pairs is Cohn et al. 2005a,b, which tested words in lists, as well as in constructed sentences, some of which were the same as in Guion's study.

It is not immediately clear why some experimental studies found homophones to differ in duration while others did not. Small sample size may be one reason: the largest sample studied experimentally was that of Whalen (1996), which included twenty-five pairs. Guion (1995) and Cohn and colleagues (2005a,b) tested four and fourteen pairs, respectively. Another problem is that experimentally observed word durations in part reflect the mechanics of experimentation itself, such as presentation order, an issue I return to in §7.1 below. What the experimental record does not show is whether homophones come to sound different when grouped by frequency, or whether they become more similar in generic carrier phrases or word lists. A way to address these questions is to look at homophone durations in spontaneous speech.

A small number of studies have examined homophone durations in corpora of speech in naturalistic settings. Lavoie 2002 examined the pronunciation of the words *four* and *for* in read speech and in spontaneous speech. Although Lavoie reported shorter durations for the more frequent *for* than for the less frequent *four*, those differences may be related to the prosodic environments of the two items in question, which affects the contextual speaking rate. Indeed, Lavoie's interpretation of the durational differences is that they reflected effects of articulation in context, rather than differences in the representation associated with each word.

Jurafsky et al. 2002 examined the durations of four ambiguous function words (*to*, *that*, *of*, and *you*) in a subset of the Switchboard corpus of American English telephone conversations. A subsequent study (Bell et al. 2003) examined the ten most frequent English function words. Using multiple regression, Bell and colleagues controlled for factors known to affect duration, such as speaking rate, segmental context, pitch accent, and contextual predictability. Once these factors were controlled for, the frequency of

---

[2] The editor points out that speakers could come to treat synchronic variation in the pronunciation of homophone pairs as meaningful, along similar lines as speakers' tendency to invest phonemically different pronunciations of one and the same word with different meanings (e.g. *vase* rhyming with *face* or *Oz* when referring to either ordinary household objects or valuable pieces of art). The words investigated by Nygaard and colleagues may be moving along such a path. The important point for the current discussion is that the pronunciation alternants in that study did not differ in phonemic content.

the preposition *to* vs. the infinitival marker *to*, for example, was no longer a significant predictor of word durations. In a review of frequency effects, Jurafsky (2003) concludes that, when other factors are held constant, there are no effects of lemma frequency on pronunciation, either in elicited productions or in spontaneous speech. However, function-word production is thought to draw on different mechanisms from those employed in content-word production in many models of language production (Bock & Griffin 2000, Garrett 1988, Levelt et al. 1999). Therefore, observations on function words may not generalize to content words.

Why does word duration vary? Several mutually compatible explanations have suggested themselves: word durations tend to shorten, and articulatory effort tends to be reduced, as a function of repetition within a discourse (Bard et al. 2000, Fowler 1988, Fowler & Housum 1987, Fowler et al. 1997, Shields & Balota 1991), predictability within an utterance (Gregory et al. 1999, Hunnicutt 1985, Jurafsky et al. 2001a,b, Lieberman 1963), and neighborhood density (Wright 2004).[3] In all of these cases, high-probability forms tend to reduce, and low-probability forms tend to lengthen or otherwise be hyperarticulated. Explanations for shortening and the twin phenomenon of lengthening have ranged from theories attributing variation to factors under speakers' control, such as a desire to provide listeners with maximally distinctive information while minimizing articulatory effort (e.g. Lindblom 1990), to theories emphasizing factors not under speakers' control, which include articulatory practice, speed of lexical access, effects of associative priming, and factors during stages of lexical production that follow lexical access (Balota et al. 1989, Shields & Balota 1991).

In sum, word durations can offer valuable clues to the mechanisms underlying language production in general and to the locus—or loci—of frequency in the lexicon in particular. Homophone durations in particular can elucidate whether lemma frequency, as distinct from form frequency, can affect duration. Since many factors affect content-word durations, these other factors need to be controlled experimentally or statistically if we are to understand whether lemma frequency affects word duration. The experimental record is difficult to interpret, since some observed pronunciation variation necessarily arises from experimental design itself. Existing corpus-based studies have so far failed to turn up differences in homophone duration when other factors were held constant, consistent with the notion that the word form is the locus of frequency in the lexicon. But previous corpus-based studies were based on small samples or on function words, which may call on processing mechanisms different from those for content words. What is needed, therefore, is an analysis of a large sample of homophonous content words in spontaneous speech. Providing such an analysis is the goal of this article.

3. METHOD. A list of all items that were homophonous with at least one other item in the CELEX database of English lexemes (Baayen et al. 1993) was automatically created. Since the subsequent corpus searches depended on an orthographic transcript of ca. three million words, the search was restricted to pairs that differed in spelling (e.g. *steak* and *stake*). Identifying lemmas as same or different raises many difficult questions, and it would not have been feasible to make this determination for all ambiguous items in this study. Therefore, all items with identical spelling were pooled together. For example, the plural noun and the third-person singular verb *laps* were treated as

---

[3] Neighborhood density refers to the size of the set of words that are phonemically similar to a given word. Effects of neighborhood density on perception have been known for a long time. More recent literature, of the past ten years or so, recognizes that neighborhood density also affects speech production (see Dell & Gordon 2003 for an overview).

a single item in the search, paired with the form *lapse* (which itself is also category-ambiguous). I return to the questions posed by category-ambiguous items in §6.1 below. In a few cases, a given pronunciation was associated with more than two items, yielding triplets such as *praise*, *prays*, and *preys*, or quadruplets such as *right*, *write*, *rite*, and *wright*. Subsequent analyses included only the two highest-frequency members of such sets. Ten words, represented by a total of sixty-two tokens, were excluded for this reason.

The analysis took into account all tokens of the homophone candidates from the Switchboard corpus of American English (Godfrey et al. 1992), a collection of 240 hours of telephone conversations between strangers. A time-aligned orthographic transcript of the corpus (Deshmukh et al. 1998) gives the beginning and end time of each word, in milliseconds, measured from the beginning of each conversation.

Several classes of homophone candidates in CELEX were excluded from the analysis. Words with nonhomophonous homographs formed one such class. For example, the noun *tear* is homophonous with *tier*, but also homographic with the nonhomophonous verb *tear*. Of the set of attested homophone pairs, twelve were excluded for this reason, based on the pronunciations given in CELEX. Another class that was excluded consisted of pairs that had identical frequencies in Switchboard.[4] In addition, in a manual search through the remaining attested homophone pairs, the following types of items were excluded: (i) pairs such as *source-sauce* that are homophones according to the British English (RP) pronunciations represented in CELEX, but that are not homophonous for most speakers of American English; (ii) items representing transcription errors in CELEX; for example, *texture* and *textured* are erroneously transcribed as having the same pronunciation; (iii) names of letters of the alphabet; and (iv) function words, such as *in* and *or* (paired with *inn* and *ore*), and interjections, such as *whoa* (paired with *woe*).

How would function words affect the results? Table 1 shows the thirteen most frequent function words with homophonous content words in Switchboard. Function words are approximately 60 milliseconds shorter on average than their content-word twins (210 ms vs. 273 ms), and they are also vastly more frequent (median token frequency 8,987 vs. 46). This need not constitute evidence for an effect of lemma frequency, however; as mentioned earlier, most theories of language production assume that the mechanisms underlying the production of function words differ from those for generating content words. To avoid inflating any observed effect of lemma frequency in the analysis, function words and their homophone twins were excluded from the sample.

For all tokens of the usable homophone pairs in the Switchboard corpus, the duration was extracted and the average duration of each form was computed. The log-transformed average durations of the high-frequency and low-frequency homophones were then compared using paired *t*-tests.

4. RESULTS: COMPARING MEAN DURATIONS. There were 223 homophone pairs meeting the criteria for inclusion, represented by 80,179 tokens. The higher-frequency members of the pairs had a mean frequency of 367.7 in the Switchboard corpus, with a median of 33. The mean Switchboard frequency of the lower-frequency forms was 21.6, with a median of 4.

---

[4] Additional analyses were performed in which pairs with identical Switchboard frequencies were included, and in which the classification into high-frequency and low-frequency members of homophone pairs was based on CELEX. The results were qualitatively identical to those reported here, both for the average durations and for the regression analysis.

| SPELLING | | FREQUENCY | | AVERAGE DURATION | |
FUNCTION WORD	CONTENT WORD	FUNCTION WORD	CONTENT WORD	FUNCTION WORD	CONTENT WORD
in	inn	39,968	13	143	274
but	butt	28,913	13	212	326
so	sew	27,145	46	264	361
we	wee	25,379	5	149	258
not	knot	14,395	1	233	200
one	won	12,657	146	212	290
some	sum	8,987	17	233	297
would	wood	8,288	192	164	281
no	know	7,980	48,002	300	180
where	wear	5,581	572	192	248
our	hour	5,072	531	168	291
which	witch	2,924	5	222	286
through	threw	1,923	57	243	250

TABLE 1. Frequency and token duration of the thirteen most frequent function words with homophone twins in the Switchboard corpus.

A comparison of the average duration of each pair's lower-frequency member to that of its higher-frequency counterpart revealed that the lower-frequency words were longer than their high-frequency counterparts (396 vs. 368 ms on average). A paired *t*-test showed this difference to be significant ($t(222) = -3.1382$, $p < 0.01$).

5. DISCUSSION: COMPARING MEAN DURATIONS. Average durations of the high-frequency words were shorter than those of their low-frequency homophones, consistent with the hypothesized frequency effect at the lemma level. However, many factors besides frequency affect word durations. While the sample was large, suggesting that random noise was unlikely to produce the overall difference in means, it is possible that extraneous factors systematically lengthened low-frequency words and shortened high-frequency words. This possibility was explored in a multiple linear regression analysis.

6. REGRESSION ANALYSIS. A regression model relates a set of predictor variables to an outcome variable. The model coefficients indicate to what extent, on average, each predictor predicts the outcome when all other factors are taken into account. The duration of the low-frequency homophones and other factors that can affect word durations were used as the predictor variables, and the duration of the high-frequency homophones constituted the outcome variable. The question is whether lemma frequency significantly improves predictions of word duration, over and above the contribution of other factors.[5]

6.1. PREDICTOR VARIABLES.
SPEAKING RATE. What other factors besides frequency might produce duration differences between low-frequency and high-frequency words? One such factor is contextual speaking rate. If low-frequency words systematically occur in contexts with low speaking rates, then one can expect them to be relatively long. Consequently, previous studies have attributed apparent effects of lemma frequency on word durations to effects of local speaking rate (e.g. Cohn et al. 2005a,b, Lavoie 2002). By contrast, if the effect

---

[5] There are alternative models that can usefully be considered. I am currently exploring a mixed-effects model that takes into account information about each token in the database. Some interesting aspects of the data are revealed only by the mixed-effects model, but they are not germane to the issue at hand. Therefore, for ease of exposition, only the model based on per-word averages is presented here.

is stable when speaking rate is controlled for, then word durations may be a function of properties of the words themselves, such as their frequency.

The place of speaking rate in models of word duration is complicated by the fact that target-word frequency may itself contribute to variation in contextual speaking rate. As a conservative step, since we wish to understand whether apparent effects of lemma frequency might be spurious, the model initially included contextual speaking rate as a predictor in the model, to give speaking rate a chance to 'explain away' the frequency effect. The resulting model is problematic because of the suspected causal relationship between target frequency and contextual speaking rate, and because of increased collinearity among predictors. The model reported here therefore excludes contextual speaking rate. The pattern of significance for the other predictors is the same in the two models.

Two measures of contextual speaking rate (as syllables/sec) were calculated for each homophone token, one for the stretch of speech preceding the homophone in an utterance, and one for the stretch of speech following the homophone, as follows: all words preceding and following the homophone token within the utterance were extracted from Switchboard, along with their syllable counts in the CELEX database. The segmentation of Switchboard used here defines an utterance as a stretch of speech by a single speaker with a maximum duration of about 10 seconds 'not interrupted by significant pause boundaries (about 0.4–0.5 seconds of silence at each end)' (Deshmukh et al. 1998:2). The duration of the regions preceding and following each homophone token was also extracted. The speaking rate of each region was the number of syllables in each region, divided by the duration of that region.[6] Although the correlation between the speaking rates before and after the homophone targets is significant, the coefficient is not high (Pearson $r = .249$), indicating that they have the potential of measuring different things, so both of these factors were included in the model.

CONTEXTUAL PREDICTABILITY. A second potentially confounding factor concerns predictability from neighboring words. For example, Jurafsky and colleagues (2002) found that lemma frequency was not a significant predictor of word duration when contextual predictability (as measured by bigram probability) was controlled for. To control for this variable, two measures of contextual predictability were calculated for each homophone token: the conditional probability of the homophone, given the previous word, and the conditional probability of the homophone, given the following word. The conditional probability of a homophone token given the preceding word was estimated as the ratio of two frequency counts: the frequency count of the string of two words ($C(w_{i-1}w_i)$), divided by the frequency of the preceding word ($C(w_{i-1})$). Analogously, the conditional probability of a homophone token given the following word was the count of the two words occurring together ($C(w_iw_{i+1})$), divided by the frequency count of the following word ($C(w_{i+1})$). Bell and colleagues (2003), comparing the contributions of local bigram probability and of predictability measures taking into account additional words before and after the target word, found that probability of wider contexts did not significantly influence word duration once local bigrams were controlled for. Therefore, wider contexts were not included in the model reported here.

---

[6] An alternative measure of speaking rate that has been used in previous studies (e.g. Bell et al. 2002, Bell et al. 2003) is the total number of syllables in an utterance, including the target word, divided by the duration of the utterance. I consider that alternative measure to be problematic, and departed from it, because it partially confounds the measure of contextual speaking rate with the duration of the target word. When entered into the regression model instead of the unconfounded measures, the confounded measure of speaking rate does not reach significance.

SYNTACTIC CATEGORY. A third factor that might have produced the difference in homophone durations arises from the interplay between prosody and syntax. Words (or more accurately, syllables) in final position within phonological phrases tend to lengthen, a phenomenon known as phrase-final lengthening (Beckman & Edwards 1990, Ferreira 1993). The probability of a word's occurring in final position within a prosodic constituent is in part a function of its syntactic category—and syntactic categories differ in token frequency. The distribution of nouns and verbs in prosodic constituents is particularly relevant here. For example, Watson and colleagues (2008) show that speakers are more likely to place intonational boundaries after nouns than after verbs (see also Nespor & Vogel 1986, Sorensen et al. 1978), which suggests that nouns may be subject to phrase-final lengthening more often than verbs. Since nouns also have a lower average frequency than do verbs, again according to the CELEX database, phrase-final lengthening may confound frequency.

To separate effects of syntactic category and frequency, an estimate was calculated of the probability with which a given word represented a noun, based on the frequency counts in CELEX. That probability (the word's NOUN PROPORTION) was then included as an additional predictor in the regression model.

PROXIMITY TO PAUSES. To probe the possible role of phrase-final lengthening more directly, ideally one would add information about phrase boundaries to the model. Unfortunately, only a portion of the Switchboard has been prosodically annotated in a way that would make this possible. As an approximation to determining phrasal position, a further predictor in the model was the proportion of tokens of each lemma that immediately preceded a pause. For the purposes of this discussion, a pause was arbitrarily defined as a silence of at least 0.5 seconds.

ORTHOGRAPHY. Another factor that may affect homophone durations is orthography. For example, Warner and colleagues (2004) report differences in the pronunciation of pairs of Dutch words with different orthography but shared underlying phonemic material. Orthography may affect homophone durations in two ways, both of which can be illustrated by the example *time* vs. *thyme*. First, *time* contains fewer letters than *thyme*, raising the possibility that a word's duration in pronunciation is related to its length in letters. Some evidence (from Dutch) that this may be so can be found in Warner et al. 2004. Second, *time* and *thyme* differ in orthographic regularity: the pronunciation of the word *time* is far more likely, given its graphemes, than the pronunciation of the word *thyme* (see Gontijo et al. 2003). More generally, predictability of a word's pronunciation, given its spelling, is higher for high-frequency words than for low-frequency words (Gontijo et al. 2003). If it is the case that irregularly spelled words tend to be longer in duration than words with regular spelling, then orthographic regularity could be responsible for the observed difference in duration.

To allow probing of both of these possibilities, the model included length in letters in the model, along with a measure of the predictability of the word's pronunciation, given its spelling. The measure used was the M-SCORE (Berndt et al. 1987), which represents the average probability of a word's graphemes, normalized by the probability of the most probable pronunciation of each grapheme. These scores for each lemma were calculated using the grapheme-to-phoneme probabilities in Berndt et al. 1987, which are based on a corpus of American English.

LEMMA FREQUENCY. The predictor variable that was crucial to the research question was the frequency of each member of a homophone pair. As is usual in research on

frequency effects, the model used log frequency, not raw frequency. This decision is motivated conceptually by the fact that the function relating word frequency to behavioral measures in many reported psycholinguistic experiments is approximately logarithmic: a given absolute difference in frequency will have a larger effect in the lower-frequency ranges than in the higher-frequency ranges (e.g. Hay 2002, Howes & Solomon 1951).

6.2. STRUCTURE OF THE MODEL. The outcome variable in the regression model was the average duration of the high-frequency homophones. To determine whether lemma frequency affects word durations when other factors are controlled for, two models were compared. The first ('baseline') model contained all factors discussed in the preceding sections, except for frequency. The second ('frequency') model contained the same factors as the baseline model, in addition to the frequency of the high-frequency homophone (i.e. of the word whose duration was being predicted). Factors were considered significant if they were at or below the 0.05 level of significance.

Classification of one or the other member of a homophone pair as the higher-frequency form was based on frequency information from Switchboard. Additional models were also constructed in which that classification was based on frequency in CELEX. The overall pattern of results is unchanged in these models, in that the same predictor variables reach statistical significance in both sets of models. The model reported here represents the model with the highest adjusted R^2 for the baseline model. This means that the results represent a conservative estimate of the effect of lemma frequency.

6.3. RESULTS OF REGRESSION ANALYSIS. Neither bigram probability given the preceding word nor speaking rate in the region preceding the target word emerged as significant predictors, in either the baseline model or the model including frequency. This result is consistent with related studies of word duration (Bell et al. 2003, Bell et al. 2008). Orthographic length also failed to reach significance in the model containing all other predictors. Therefore, a new model was fitted that did not include bigram probability given the preceding word, speaking rate in the region preceding the target, or orthographic length. The pattern of significance of all other factors remained the same in the models with and without the nonsignificant predictors. The model reported here excludes the nonsignificant factors.

The baseline model accounted for 43% of the variability in the duration of the high-frequency homophone twin. The frequency model, which differed from the baseline only in including frequency as an additional predictor, accounted for 48% of the variability. Crucially, lemma frequency emerged as a significant predictor when this factor was included in the model. An ANOVA comparing the baseline model and the frequency-based model showed that inclusion of frequency was justified ($F(1,212) = 19.65$, $p < 0.001$).

The pairwise correlations among the predictor variables are shown in Table 2. A cluster analysis of these correlations is shown in Figure 2. The speaking rate following the target word and the proportion of targets immediately followed by pauses are negatively correlated (Spearman's rho $= -.357$), reflecting the fact that, other things being equal, pausing results in lower speaking rates on average (unless speakers speed up their productions between pauses). The probability given the following word and the frequency of the target were also clustered together, but the correlation was sufficiently

low that these two variables have the potential to explain different things (Spearman's rho = .364). To minimize collinearity among predictor variables, one might wish to include only one variable from each of these clusters in the model; since the aim was to study the contribution of word frequency when other factors were taken into account, however, all variables were retained in the model. To explore the stability of the model in the face of this concern, the variance inflation factors (VIFs) were calculated, which indicate the extent to which the correlation of a given variable with other variables in the model inflates the standard error of the regression coefficient for that variable. The VIFs are shown in Table 3, along with the regression coefficients. All VIFs were below 1.4.

	high-fq duration[a]	low-fq duration[b]	m-score[c]	noun-proportion[d]	speaking rate[e]	bigram probability[f]	pauses[g]
low-fq duration[b]	0.415						
m-score[c]	−0.238	−0.177					
noun-proportion[d]	0.295	0.107	−0.036				
speaking rate[e]	−0.231	0.002	−0.051	−0.22			
bigram probability[f]	−0.338	−0.099	0.078	−0.041	−0.149		
pauses[g]	0.194	−0.029	−0.075	0.172	−0.379	0.048	
log frequency[h]	−0.417	−0.175	0.046	−0.092	0.084	0.364	0.315

TABLE 2. Spearman correlation matrix of factors affecting word duration, based on 220 homophone pairs ($N = 220$ for all correlations).

[a] Log-transformed duration of the higher-frequency member of a homophone pair.
[b] Log-transformed duration of the lower-frequency member of a homophone pair.
[c] Grapheme-phoneme probability (see text).
[d] Proportion of tokens representing nouns.
[e] Speaking rate in the region following the target word within an utterance.
[f] Conditional probability of the target word, given the following word.
[g] Proportion of tokens immediately preceding pauses.
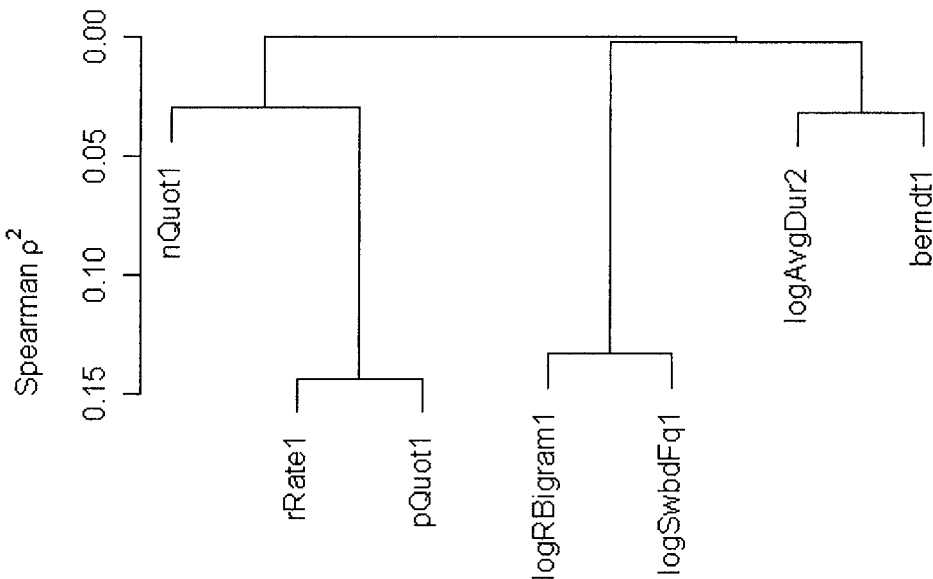[h] Log-transformed word frequency in Switchboard.

FIGURE 2. Hierarchical clustering of predictors in a regression model of homophone durations ($N = 220$).

VARIABLE	B	β	SE	t	VIF
intercept	−0.5247		0.103497	−5.07	
low-fq duration[b]	0.2141	0.2823	0.039524	5.416	1.1004
m-score[c]	−0.2213	−0.1565	0.073207	−3.023	1.0847
noun proportion	0.1034	0.2178	0.024098	4.292	1.0427
speaking rate[f]	−0.0492	−0.1386	0.020312	−2.422	1.3258
bigram probability[h]	−0.0171	−0.1826	0.005315	−3.21	1.3104
pauses[g]	0.2813	0.1187	0.136587	2.06	1.3447
log frequency[h]	−0.0297	−0.2471	0.00669	−4.433	1.2581

TABLE 3. Summary of regression model of durations of high-frequency homophones ($N = 220$); B = raw unstandardized coefficient, β = standardized coefficient, SE = standard error, t = t value, VIF = variance inflation factor.

To what extent might the model overfit the dataset? Bootstrap validation was used to obtain a corrected R^2 to learn the extent to which the model parameters are estimated to change when the model is based on a different sample. Simulations with 200 bootstrap runs yielded a corrected R^2 of .43, indicating a modest shrinkage of .05 compared to the uncorrected R^2. The only predictor that was retained in all 200 bootstrap runs was the duration of the homophone twins. The frequency predictor was retained 197 times. The only other factor that was retained as often as frequency was the proportion of noun uses, one of the proxy measures for phrase-final lengthening. Bigram probability and orthographic regularity were also in most of the models (191 and 182 times, respectively). Speaking rate and proportion of prepausal tokens were retained in the majority of models as well (157 and 151 times, respectively). The most dispensable predictor was length in letters, which was retained in only 89 runs. This pattern is consistent with the behavior of these factors in other models of the dataset.

A striking aspect of the model is the small contribution of homophone duration as a predictor of word duration. Homophones are usually defined as sets of words that sound alike. Given that definition, one would expect the duration of a word like *thyme* to predict the duration of its twin *time* perfectly. That is not the case. A model containing homophone duration as the sole predictor accounts for just 19% of the variability in duration. It is clear that other factors besides a word's phonemic makeup influence word duration to a considerable degree. As Table 3 shows, grapheme-phoneme probability (m-scores), the estimated proportion of noun tokens of an orthographic word (the word's 'noun proportion'), speaking rate in the region following the target word, the conditional probability of the target word given the following word, and the proportion of tokens immediately preceding pauses all predicted target duration, in the hypothesized manner: high m-scores, fast speaking rate, and high bigram probability all predict shorter durations, and high noun proportion and high proportion of prepausal tokens predict longer durations. Each of these factors is individually significant when all other factors are in the model, as revealed by a nonsequential ANOVA.

Crucially for the current study, the log frequency of a word was a significant predictor of word duration when all other factors were controlled for: as frequency increases, word duration decreases, when other factors are held constant. This effect, while small, is similar in size to other theoretically important effects on word duration reported in the literature, such as effects of repetition, associative priming, and contextual predictability (e.g. Bell et al. 2003, Shields & Balota 1991), and to the effects of the other factors in the model.

7. DISCUSSION OF REGRESSION MODEL: EFFECTS OF REPETITION AND CHOICE OF OUTCOME VARIABLE. The regression model suggests that lemma frequency affects word duration

above and beyond other factors. However, words shorten if they occur multiple times within a discourse (Fowler 1988, Fowler & Housum 1987, Fowler et al. 1997). This raises the possibility that the observed effect of frequency reflects repetition within a conversation, rather than overall lemma frequency. To investigate this possibility, the regression analysis was repeated, this time using only the first occurrence of a given homophone within a conversation. The first occurrences of high-frequency homophones were significantly shorter than their low-frequency counterparts (374.0 vs. 395.7 ms, $t(228) = -2.981, p < 0.01$). The new model was qualitatively identical to the model based on all tokens: all factors that were significant predictors in the previous model were also significant in the model without the repeated tokens. These results suggest that the effect of overall usage frequency on word duration is separate and distinguishable from the effect of repetition within a discourse: even the first time a word occurs in a conversation, its duration in part reflects its frequency.

8. GENERAL DISCUSSION. The central finding of this study is that homophone pairs in a corpus of spontaneous speech differed in duration, with high-frequency words being shorter than their lower-frequency homophone twins. The effect of frequency on word duration remained significant when local speaking rate, syntactic category, predictability from neighboring words, proximity to pauses, length in letters, and orthography were statistically controlled. These results yield two theoretical implications: first, they suggest that lemma frequency affects language production, contra Levelt and colleagues (1999) and Jurafsky (2003). Second, they suggest that the shorter duration of frequent forms cannot solely be due to increased articulatory routinization with highly practiced items, as suggested by Bybee (2001) and Newmeyer (2006). It may be possible to account for these results in models in which lexical representations include subphonemic detail (e.g. Johnson 1997, Pierrehumbert 2001, 2002).

I begin the discussion by examining these results in the context of previous experimental and corpus studies of homophone durations. I then consider how these results may be reconciled with seemingly contradictory previous findings, particularly effects of form frequency. Finally, I turn to further theoretical implications of these results, concerning the nature of phonological representations and the sources of shortening of frequent forms.

8.1. HOMOPHONE DURATIONS IN SPONTANEOUS SPEECH AND IN THE LAB. Could the observed durational differences be due to uncontrolled variation? Using a corpus of spontaneous speech makes it impossible to control for all factors that affect duration. Two factors in particular pose potential problems: conceptual familiarity and givenness in a discourse. Controlled experiments show that conceptualization takes longer for concepts with low-frequency names (Griffin & Bock 1998, Levelt et al. 1999). But the experimental record suggests that low-frequency words are longer in duration than high-frequency words even when conceptual familiarity is controlled (Bonin & Fayol 2002). Also, when low-frequency words do occur in the corpus, their occurrence is motivated by the context, which primes relevant concepts and thus may facilitate conceptualization. Therefore, effects of conceptual familiarity may be smaller in the corpus than in experiments involving unprimed single-word production. In other words, conceptual familiarity is unlikely to be responsible for the observed duration differences.

A second factor that was impossible to control for in the current study concerns discourse context, which affects accent placement and other aspects of pronunciation (Fowler 1988, Fowler & Housum 1987, Fowler et al. 1997, Terken & Hirschberg 1994).

In this connection, it is interesting to note that Guion (1995) found durational differences when homophones were embedded in full sentences, but not in generic carrier phrases. This is consistent with a pattern in which experimental stimuli in meaningful contexts display variation resembling that found in the corpus, whereas stimuli in generic carrier phrases pattern more like word lists, which do not reliably give rise to differences in homophone durations. The fact that such differences have been observed for single-word production (Bonin & Fayol 2002, Whalen 1996) suggests that it cannot be entirely due to accent placement or other aspects of language use in particular contexts.

Given that homophones differ in duration in our corpus of spontaneous speech, is it possible to explain why experimental studies have not consistently found differences in homophone durations? One possible explanation relates to presentation order. As noted earlier, one study (Whalen 1996) found differences when homophones were presented in word lists grouped by frequency, but not in mixed lists. Such an effect of stimulus blocking is expected if lemma frequency affects durations: findings by Kello and Plaut (2000, 2003) suggest an explanation for why frequency effects would be less likely to surface in mixed word lists than in lists grouped by frequency. Kello and Plaut show that subjects tend to read word lists at a regular pace, in effect setting themselves a deadline for each item. For lists with high-frequency items, which subjects are able to produce faster, that deadline is faster than for low-frequency items. In mixed lists, subjects set themselves a generic deadline, which attenuates the effects of properties of individual stimuli. The finding by Bonin and Fayol (2002) of faster picture-naming latencies for high-frequency compared to low-frequency homophones in French may lend itself to a similar explanation: Bonin and Fayol employed a between-subjects design, in which some subjects saw only high-frequency items, and others only low-frequency ones. Perhaps subjects in the picture-naming task tended to pace themselves in a manner similar to that described by Kello and Plaut. If this is so, then one would expect stronger effects of frequency when subjects see only high-frequency or only low-frequency items than when stimuli are presented in mixed lists.

When one considers corpus evidence and experimental evidence together, it becomes clear that differences in homophone duration in the corpus cannot solely be due to these uncontrolled modulating factors, since they have also emerged when modulating factors were controlled experimentally. Conversely, the fact that homophone pairs differ in duration in naturalistic settings suggests that this effect is not entirely attributable to the mechanics of experimentation, such as presentation order. At the methodological level, then, our findings illustrate that corpora of spontaneous speech and controlled experiments are two one-legged creatures that can usefully support one another.

8.2. Differences in homophone duration and effects of form frequency. The present findings pose a challenge to the claim made in Levelt et al. 1999 that word frequency is a property of the phonological form, and not of the lemma. Recall that this claim was motivated by the phenomenon of frequency inheritance, that is, the observation that low-frequency words with high-frequency homophones behave in some respects as though they were of high frequency (Dell 1990, Jescheniak & Levelt 1994). There are at least three ways to reconcile the present findings with these seemingly contradictory observations. The first is to note that frequency inheritance may in fact have affected word durations in the corpus: the low-frequency words were longer than their high-frequency twins, but perhaps they were shorter than they would be without those twins. Unlike Jescheniak and Levelt's experiments, this corpus-based study did not include words without homophones as a control condition.

A second way to reconcile the present findings with previously observed effects of form frequency relates to spelling: it is possible that frequency inheritance is stronger for homophones that are spelled alike than for homophones that differ in spelling. Jescheniak and Levelt's findings were based on homographic homophones, whereas the present study was based on homophones that differed in spelling. As Bonin and Fayol (2002) point out, if there is feedback from the orthography of a word to the conceptual level, or from orthography to phonological form, then low-frequency words with high-frequency homographs should have a processing advantage. In this connection, it is interesting to note that the effect of lemma frequency observed by Bonin and Fayol was stronger in written than in spoken naming latencies. It is plausible to suppose that orthography affects writing more strongly than speaking.

A third way to reconcile the present findings with those of Jescheniak and Levelt relates to the persistence of the lemma-frequency effect. Along with a robust effect of phonological form, Jescheniak and Levelt (1994) found a transient effect of lemma frequency. This raises the question of whether the effect holds ONLY for the first occurrence of a word within a discourse. Bonin and Fayol (2002) found that spoken and written naming latencies were reliably different for low- vs. high-frequency homophones, and that the differences persisted across four repetitions. Does the effect on durations persist as well? The fact that lemma frequency emerged as a significant predictor of duration regardless of whether repeated tokens were included in the regression model suggests that the effect might persist, but does not prove it. Unfortunately, it would not be feasible to take into account, say, only the fourth occurrences of words in a given conversation in the Switchboard corpus: the conversations in Switchboard are short, so repeated productions of low-frequency words within a conversation are naturally rare. A related problem is that repetition within a discourse also induces shortening. The question then becomes whether words are longer than their high-frequency twins even when they are repeated. Preliminary results (Bell et al. 2008) indicate that repetition and frequency each affect word duration independently. But since that study, unlike the current study, did not control for form frequency, it does not elucidate whether the relevant frequency measure is a property of forms or of lemmas. The persistence of lemma-frequency effects on word duration needs to be investigated further.

8.3. CONVERGING EVIDENCE ON LEMMA-FREQUENCY EFFECTS. At least two other recent studies have challenged the notion that the phonological form is the only locus of frequency in word production. One study (Navarrete et al. 2006) finds that lemma frequency affects latencies in a gender-decision task and in the production of gender-marked pronouns. Since these tasks require lemma retrieval, but not retrieval of a phonological form, such effects run counter to the predictions of models in which frequency is a property of phonological form. It should be noted that Navarrete and colleagues' observations are based on Spanish. As mentioned above, Jescheniak and colleagues (2003) suggest that frequency inheritance could be stronger in some languages than in others. It is noteworthy, therefore, that the present study found lemma-frequency effects in English, a language for which frequency inheritance is well established.

Another recent study (Kittredge et al. 2008) lends further support to the notion that lemma frequency affects lexical access in English. The evidence in that study comes from aphasic naming errors. Earlier studies of single-word production in aphasia had concluded that frequency affects phonological errors, not semantic ones (an exception

is Nickels & Howard 1994, cited in Kittredge et al. 2008). Kittredge and colleagues demonstrate that semantic errors reflect frequency sensitivity in a group of thirty-one individuals with various aphasic syndromes. Kittredge and colleagues describe how effects of lemma frequency may be accounted for in a computational model (Schwartz et al. 2006). Frequency, on that model, does not reside exclusively at the level of phonological form, but pervades the model's network.

8.4. DIFFERENCES IN HOMOPHONE DURATION AND PHONOLOGICAL REPRESENTATIONS. One set of implications of these results relates to the nature of phonological forms. The models of language production mentioned so far assume that phonological representations are composed of an alphabet of discrete segments. Subphonemic detail, on that view, is a matter of phonetic implementation, not phonological representation. This assumption is shared with most models of auditory word recognition (e.g. McClelland & Elman 1986, Norris 1994). This segment-based conception of lexical representations has increasingly come to be challenged by work suggesting that lexical representations include subtle phonetic detail, and that word recognition relies on such detail (see e.g. Johnson 2004, Kemps et al. 2005, Salverda et al. 2003). In a phoneme-based representation, *thyme* and *time* are identical. Segment-based models usually account for systematic differences between items with identical phonemic content by attributing such differences to variations in phonetic implementation, for example, variations in local speaking rate (e.g. Lavoie 2002; see also Levelt 1989:Ch. 9). As we saw earlier, such a move is problematic for the current data, since homophones still differed in duration when local speaking rate and other factors were statistically controlled.

An alternative conception of phonological representation is available in a family of models in which representations of segments, words, and phrases include fine phonetic detail, and in which a given word potentially has large numbers of representations. In these so-called exemplar-based models, memory representations of utterances are stored in the mind as separate exemplars, and these exemplars are activated during both the production and the perception of speech (Bybee 2001, Gahl & Yu 2006, Goldinger 1997, 1998, Hay & Bresnan 2008, Johnson 1997, 2004, Pierrehumbert 2001, 2002). One and the same word can be represented by many exemplars. In principle, there can be as many exemplars as tokens in a speaker's experience. In actuality, exemplar-based models usually assume that the number of exemplars stored is smaller than the number of tokens in experience: the number of exemplars is also affected by memory decay and by abstracting over tokens, that is, mapping multiple tokens onto single exemplars. The information associated with an exemplar is not restricted to phonetic detail. It may also include information about the situational context of the utterance, as well as information about the identity, gender, age, perceived social status of the speaker, and so on.[7] Different tokens of one and the same item will by necessity differ in one way or another. Exemplar-based conceptions of the lexicon are well poised to model

---

[7] A referee raises the question of why some factors, such as speakers' gender and age, are registered in exemplar-based models, while other factors, such as time of day or temperature, are not registered, or are never linguistically relevant. The general answer to this question is that, during early stages of language development, all factors may be registered. Over time, the attentional space in which categorization takes place adjusts in such a way that factors come to affect categorization to a greater or lesser degree depending on the degree to which they have been found to affect linguistic meaning. These ideas are rooted in psychological models of categorization and attention (Nosofsky 1986) and have been incorporated into exemplar-based models of language learning (Regier 2005).

frequency effects. Frequent items give rise to a greater number of exemplars, or else to exemplars that have been activated more frequently.

The shortening of frequent forms can be conceptualized as follows in an exemplar-based model. Compared to low-frequency words, high-frequency words are more likely to shorten due to a range of factors: high-frequency words are accessed faster, tend to be more predictable in discourse, elicit weaker articulatory effort, may benefit from articulatory routinization, and so on. All of these factors conspire to change the distribution of exemplars representing high-frequency lemmas so as to include a relatively greater number of short tokens—and thus shorten the average duration. This change in the distribution leads to a gradual change in articulatory target, a process that contributes to further shortening: subsequent productions are increasingly likely to be short.

8.5. Differences in homophone duration and the place of frequency in the lexicon. The finding that shortening takes into account information about the identity of a lemma calls into question the suggestion (made, for example, in Newmeyer 2006) that the shortening of frequent forms is analogous to practice effects in motor skills or neuromotor fluency. As mentioned in the introduction, Bybee (2001) similarly cites 'neuromotor fluency' as the mechanism for reductive change. The present findings suggest that form frequency, or the frequency of particular combinations of segments, is insufficient for predicting which forms shorten. Instead, lemma frequency, that is, frequency indexed by information about a word's meaning and syntactic properties, is a determinant of word duration. This means that the shortening of frequent forms is not purely the result of increased motor fluency.

The notion that frequent forms shorten as a result of increased motor fluency has been shared across a wide spectrum of opinions. As unfortunate as it may seem that this widely shared notion turns out to be untenable, the hope is that recognizing the limited role of articulatory fluency in shortening of frequent forms will aid an increased understanding of the relationship between language usage and linguistic representation.

Despite the emphasis in some usage-based accounts (such as Bybee 2001) on articulatory routinization, it is clear that that work is in fact consistent with the findings presented here: for example, a number of such accounts (Bybee 2002a,b) clearly entail that reduction processes are word-specific and context-specific. More fundamentally, the usage-based work of Bybee and others shares with the current work the view that frequency shapes linguistic knowledge profoundly and affects all aspects of language production and comprehension.

9. Conclusion. One motivation for Levelt and colleagues' (1999) decision to propose the phonological form as the sole locus of frequency information in the lexicon was parsimony. On the face of it, a model that includes only one locus of frequency information appears simpler than one that includes multiple loci for such information. However, I agree with the observation that 'parsimony cannot be assumed to be a property of the language system; it is only something to which accounts of its underlying principles aspire' (O'Seaghdha 1999:51). The underlying principle of recognizing that frequency may shape every aspect of language and speech is simple.

REFERENCES

Baayen, R. Harald; Richard Piepenbrock; and H. van Rijn. 1993. *The CELEX lexical database (CD-ROM)*. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

BALOTA, DAVID A.; JULIE E. BOLAND; and LYNNE W. SHIELDS. 1989. Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory and Language* 28.14–36.

BARD, ELLEN GURMAN; ANNE H. ANDERSON; CATHERINE SOTILLO; MATTHEW AYLETT; GWYNETH DOHERTY-SNEDDON; and ALISON NEWLANDS. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language* 42.1–22.

BECKMAN, MARY E., and JAN EDWARDS. 1990. Lengthenings and shortenings and the nature of prosodic constituency. *Papers in laboratory phonology 1: Between the grammar and the physics of speech*, ed. by John Kingston and Mary E. Beckman, 152–78. Cambridge: Cambridge University Press.

BELL, ALAN; JASON BRENIER; MICHELLE GREGORY; CYNTHIA GIRAND; and DANIEL JURAFSKY. 2008. Predictability effects on content versus function word durations in conversational English. *Journal of Memory and Language*, to appear.

BELL, ALAN; MICHELLE L. GREGORY; JASON BRENIER; DANIEL JURAFSKY; AYAKO IKENO; and CYNTHIA GIRAND. 2002. Which predictability measures affect content word duration? Paper presented at Pronunciation Modeling and Lexicon Adaptation for Spoken Language Technology Workshop, Estes Park, CO.

BELL, ALAN; DANIEL JURAFSKY; ERIC FOSLER-LUSSIER; CYNTHIA GIRAND; MICHELLE GREGORY; and DANIEL GILDEA. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America* 113.1001–24.

BERNDT, RITA S.; JAMES A. REGGIA; and CHARLOTTE C. MITCHUM. 1987. Empirically derived probabilities for grapheme-to-phoneme correspondences in English. *Behavior Research Methods, Instruments & Computers* 19.1–9.

BOCK, J. KATHRYN. 1995. Sentence production: From mind to mouth. *Speech, language, and communication*, ed. by Joanne L. Miller and Peter D. Eimas, 181–216. San Diego: Academic Press.

BOCK, J. KATHRYN, and ZENZI M. GRIFFIN. 2000. Producing words: How mind meets mouth. *Aspects of language production*, ed. by Linda Wheeldon, 7–47. New York: Psychology Press.

BONIN, PATRICK, and MICHEL FAYOL. 2002. Frequency effects in the written and spoken production of homophonic picture names. *European Journal of Cognitive Psychology* 14.289–314.

BYBEE, JOAN. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

BYBEE, JOAN. 2002a. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24.215–22.

BYBEE, JOAN. 2002b. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14.261–90.

BYBEE, JOAN. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–33.

CARAMAZZA, ALFONSO. 1997. How many levels of processing are there in lexical access? *Cognitive Neuropsychology* 14.177–208.

CARAMAZZA, ALFONSO; ALBERT COSTA; MICHELE MIOZZO; and YANCHAO BI. 2001. The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27.1430–50.

COHN, ABBY; JOHANNA BRUGMAN; CLIFFORD CRAWFORD; and ANDREW JOSEPH. 2005a. Lexical frequency effects and phonetic duration of English homophones: An acoustic study. *Journal of the Acoustical Society of America* 118.2036.

COHN, ABBY; JOHANNA BRUGMAN; CLIFFORD CRAWFORD; and ANDREW JOSEPH. 2005b. Phonetic durations of English homophones: An investigation of lexical frequency effects. Paper presented at the annual meeting of the Linguistic Society of America, Oakland, CA.

DELL, GARY S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93.283–321.

DELL, GARY S. 1990. Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes* 5.313–49.

DELL, GARY S., and JEAN K. GORDON. 2003. Neighbors in the lexicon: Friends or foes? *Phonetics and phonology in language comprehension and production*, ed. by Niels O. Schiller and Antje S. Meyer, 9–47. New York: Mouton de Gruyter.

DESHMUKH, NEERAJ; ARAVIND GANAPATHIRAJU; ANDI GLEESON; JONATHAN HAMAKER; and JOSEPH PICONE. 1998. Resegmentation of Switchboard. Paper presented at the International Conference on Spoken Language Processing, Sydney, Australia.

FERREIRA, FERNANDA. 1993. Creation of prosody during sentence production. *Psychological Review* 100.233–53.

FERREIRA, VICTOR S., and ZENZI M. GRIFFIN. 2003. Phonological influences on lexical (mis)-selection. *Psychological Science* 14.86–90.

FOSLER-LUSSIER, ERIC, and NELSON MORGAN. 1999. Effects of speaking rate and word predictability on conversational pronunciations. *Speech Communication* 29.137–58.

FOWLER, CAROL A. 1988. Differential shortening of repeated content words produced in various communicative contexts. *Language & Speech* 31.307–19.

FOWLER, CAROL A., and JONATHAN HOUSUM. 1987. Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory & Language* 26.489–504.

FOWLER, CAROL A.; ELENA T. LEVY; and JULIE M. BROWN. 1997. Reductions of spoken words in certain discourse contexts. *Journal of Memory & Language* 37.24–40.

GAHL, SUSANNE, and ALAN C. L. YU (eds.) 2006. *The Linguistic Review* 23.3. (Special issue on Exemplar-based models in linguistics.)

GARRETT, MERRILL F. 1988. Processes in language production. *Linguistics: The Cambridge survey, vol. 3. Language: Psychological and biological aspects*, ed. by Frederick J. Newmeyer, 69–96. Cambridge: Cambridge University Press.

GODFREY, JOHN J.; EDWARD C. HOLLIMAN; and JANE McDANIEL. 1992. Switchboard: Telephone speech corpus for research and development. Paper presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), San Francisco, CA.

GOLDINGER, STEPHEN D. 1997. Words and voices: Perception and production in an episodic lexicon. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullenix, 33–66. London: Academic Press.

GOLDINGER, STEPHEN D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105.251–79.

GONTIJO, POSSIDONIA F. D.; ISA GONTIJO; and RICHARD SHILLCOCK. 2003. Grapheme-phoneme probabilities in British English. *Behavior Research Methods, Instruments & Computers* 35.136–57.

GREGORY, MICHELLE L.; WILLIAM D. RAYMOND; ALAN BELL; ERIC FOSLER-LUSSIER; and DANIEL JURAFSKY. 1999. The effects of collocational strength and contextual probability in lexical production. *Chicago Linguistic Society* 35.151–66.

GRIFFIN, ZENZI M. 2002. Recency effects for meaning and form in word selection. *Brain and Language* 80.465–87.

GRIFFIN, ZENZI M., and J. KATHRYN BOCK. 1998. Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language* 38.313–38.

GUION, SUSAN. 1995. Word frequency effects among homonyms. *Texas Linguistic Forum* 35.103–16.

HAY, JENNIFER. 2002. From speech perception to morphology: Affix ordering revisited. *Language* 78.527–55.

HAY, JENNIFER, and JOAN BRESNAN. 2008. Spoken syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review*, to appear.

HOWES, DAVIS H., and R. L. SOLOMON. 1951. Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology* 41.401–10.

HUNNICUTT, SHARON. 1985. Intelligibility vs. redundancy—conditions of dependency. *Language and Speech* 28.47–56.

JESCHENIAK, JORG D., and WILLEM J. M. LEVELT. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20.824–43.

JESCHENIAK, JORG D.; ANTJE S. MEYER; and WILLEM J. M. LEVELT. 2003. Specific-word frequency is not all that counts in speech production: Comments on Caramazza, Costa, et al. (2001) and new experimental data. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 29.432–38.

JOHNSON, KEITH. 1997. Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 145–65. San Diego: Academic Press.

JOHNSON, KEITH. 2004. Massive reduction in conversational American English. *Spontaneous speech: Data and analysis (Proceedings of the 1st session of the 10th international symposium)*, ed. by Kiyoko Yoneyama and K. Maekawa, 29–54. Tokyo: The National International Institute for Japanese Language.

JURAFSKY, DANIEL. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics*, ed. by Rens Bod, Jennifer Hay, and Stefanie Jannedy, 39–95. Cambridge, MA: MIT Press.

JURAFSKY, DANIEL; ALAN BELL; and CYNTHIA GIRAND. 2002. The role of the lemma in form variation. *Laboratory phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 1–34. Berlin: Mouton de Gruyter.

JURAFSKY, DANIEL; ALAN BELL; MICHELLE L. GREGORY; and WILLIAM D. RAYMOND. 2001a. The effect of language model probability on pronunciation reduction. Paper presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2001, Salt Lake City, UT.

JURAFSKY, DANIEL; ALAN BELL; MICHELLE L. GREGORY; and WILLIAM D. RAYMOND. 2001b. Probabilistic relations between words: Evidence from reduction in lexical production. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 229–54. Amsterdam: John Benjamins.

KELLO, CHRISTOPHER T., and DAVID C. PLAUT. 2000. Strategic control in word reading: Evidence from speeded responding in the tempo-naming task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26.719–50.

KELLO, CHRISTOPHER T., and DAVID C. PLAUT. 2003. Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language* 48.207–32.

KEMPS, RACHEL J. J. K.; LEE H. WURM; MIRJAM ERNESTUS; ROBERT SCHREUDER; and R. HARALD BAAYEN. 2005. Prosodic cues for morphological complexity in Dutch and English. *Language and Cognitive Processes* 20.43–73.

KITTREDGE, AUDREY K.; GARY S. DELL; and MYRNA F. SCHWARTZ. 2008. Aphasic picture-naming errors reveal the influence of lexical variables on production. *Brain and Language*, to appear.

KRUG, MANFRED. 1998. String frequency. *Journal of English Linguistics* 26.286–320.

LAVOIE, LISA. 2002. Some influences on the realization of *for* and *four* in American English. *Journal of the International Phonetic Association* 32.175–202.

LEVELT, WILLEM J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

LEVELT, WILLEM J. M.; ARDI ROELOFS; and ANTJE S. MEYER. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–75.

LIEBERMAN, PHILIP. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6.172–87.

LINDBLOM, BJÖRN. 1990. Explaining phonetic variation: A sketch of the H&H theory. *Speech production and speech modeling*, ed. by William J. Hardcastle and Alain Marchal, 403–39. Dordrecht: Kluwer.

LOSIEWICZ, BETH L. 1995. Word frequency effects on the acoustic duration of morphemes. *Journal of the Acoustical Society of America* 97.3243.

MCCLELLAND, JAMES L., and JEFFREY L. ELMAN. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18.1–86.

MIOZZO, MICHELE, and ALFONSO CARAMAZZA. 2003. When more is less: A counterintuitive effect of distractor frequency in the picture-word interference paradigm. *Journal of Experimental Psychology: General* 132.228–52.

MIOZZO, MICHELE, and ALFONSO CARAMAZZA. 2005. The representation of homophones: Evidence from the distractor-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31.1360–71.

MUNSON, BENJAMIN. 2001. Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research* 44.778–92.

NAVARRETE, EDUARDO; BENEDETTA BASAGNI; F.-XAVIER ALARIO; and ALBERT COSTA. 2006. Does word frequency affect lexical selection in speech production? *Quarterly Journal of Experimental Psychology* 59.1681–90.

NESPOR, MARINA, and IRENE VOGEL. 1986. *Prosodic phonology*. Dordrecht: Foris.

NEWMEYER, FREDERICK J. 2006. On Gahl and Garnsey on grammar and usage. *Language* 82.399–404.

NICKELS, LYNDSEY, and DAVID HOWARD. 1994. A frequent occurrence? Factors affecting the production of semantic errors in aphasic naming. *Cognitive Neuropsychology* 11.289–320.

NORRIS, DENNIS. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition* 52.189–234.

NOSOFSKY, ROBERT M. 1986. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115.39–57.

NYGAARD, LYNNE C.; NEELAM PATEL; and JENNIFER S. QUEEN. 2002. The link between prosody and meaning in the production of emotional homophones. *Journal of the Acoustical Society of America* 112.2444.

O'SEAGHDHA, PADRAIG. 1999. Parsimonious feedback. (Commentary on Levelt et al. 1999.) *Behavioral and Brain Sciences* 22.51–52.

PHILLIPS, BETTY S. 2000. Fast words, slow words. *American Speech* 75.414–16.

PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. *Frequency and the emergence of linguistic structure*, ed. by Joan Bybee and Paul Hopper, 137–57. Amsterdam: John Benjamins.

PIERREHUMBERT, JANET B. 2002. Word-specific phonetics. *Laboratory phonology 7*, ed. by Carlos Gussenhoven and Natasha Warner, 101–40. Berlin: Mouton de Gruyter.

PLUYMAEKERS, MARK; MIRJAM ERNESTUS; and R. HARALD BAAYEN. 2005a. Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118.2561–69.

PLUYMAEKERS, MARK; MIRJAM ERNESTUS; and R. HARALD BAAYEN. 2005b. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62.146–59.

REGIER, TERRY. 2005. The emergence of words: Attentional learning in form and meaning. *Cognitive Science* 29.819–65.

SALVERDA, ANNE PIER; DELPHINE DAHAN; and JAMES M. MCQUEEN. 2003. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition* 90.51–89.

SCHWARTZ, MYRNA F.; GARY S. DELL; NADINE MARTIN; SUSANNE GAHL; and PAULA SOBEL. 2006. A case-series test of the interactive two-step model of lexical access: Evidence from picture naming. *Journal of Memory and Language* 54.228–64.

SHIELDS, LYNNE W., and DAVID A. BALOTA. 1991. Repetition and associative context effects in speech production. *Language and Speech* 34.47–55.

SORENSEN, JOHN M.; WILLIAM E. COOPER; and JEANNE M. PACCIA. 1978. Speech timing of grammatical categories. *Cognition* 6.135–53.

TERKEN, JACQUES, and JULIA HIRSCHBERG. 1994. Deaccentuation of words representing 'given' information: Effects of persistence of grammatical function and surface position. *Language and Speech* 37.125–45.

WALSH, THOMAS, and FRANK PARKER. 1983. The duration of morphemic and nonmorphemic /s/ in English. *Journal of Phonetics* 11.201–6.

WARNER, NATASHA; ALLARD JONGMAN; JOAN SERENO; and RACHEL KEMPS. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32.251–76.

WATSON, DUANE; M. BREEN; and EDWARD A. GIBSON. 2008. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, to appear.

WHALEN, DOUGLAS H. 1991. Infrequent words are longer in duration than frequent words. *Journal of the Acoustical Society of America* 90.2311.

WHALEN, DOUGLAS H. 1996. Effects of word frequency on spoken word duration. New Haven, CT: Haskins Laboratories, MS.

WRIGHT, RICHARD. 2004. Factors of lexical competition in vowel articulation. *Papers in laboratory phonology 6*, ed. by John Local, Richard Ogden, and Rosalind Temple, 26–50. Cambridge: Cambridge University Press.

Department of Linguistics
University of California, Berkeley
1203 Dwinelle Hall
Berkeley, CA 94720-2650
[gahl@berkeley.edu]