Sound Change without Frequency Effects: Ramifications for Phonological Theory

Meredith Tamminga
University of Pennsylvania

1. Introduction

This paper addresses the relationship between lexical frequency and sound change with a view to its implications for the nature of the phonology. In generative models of phonology, frequency effects in pronunciation or language change are generally relegated to a processing component beyond the scope of phonological knowledge proper. This approach stands in contrast to exemplar theory, in which speakers' mental representations of linguistic objects are not abstract but rather comprise memory traces of fine-grained phonetic detail from their experiences with language use. 1 These exemplars form clouds that provide the target (by averaging or sampling) for new production instances. Because exemplar clouds are generally assumed to be word-based, individual words are free to diverge in their phonetic targets by differential accumulation of experiences. A common potentiating factor in such differentiation is frequency, putting frequency effects in phonetic and phonological variation at the heart of the case for exemplar theory. As Johnson puts it, “the frequency distribution of variants is part of the representation of the word; thus, the representation needs to change very little to support a sound change” (2007:30). But the ease with which exemplar theory accommodates what Pierrehumbert terms “word-specific phonetics” (2002) is also a liability for the theory. In this paper I present a case study of a sound change which is not, as exemplar theory predicts, differentiated across homonyms of different frequencies. The challenge for exemplar theory, then, is to explain why sometimes frequency effects do not arise.

One difference in use that is known to be sensitive to frequency is phonetic reduction, such as consonant lenition or vowel centralization. Reductive sound change, then, is hypothesized by exemplar theoretic phonologists to proceed more quickly in more frequent words. Bybee describes the mechanism as follows:

Given a tendency for reduction during production, the phonetic representation of a word will gradually accrue more exemplars that are reduced, and these exemplars will become more likely to be chosen for production, where they may undergo further reduction, gradually moving the words of the language in a consistent direction. The more frequent words will have more chances to undergo online reduction and thus will change more rapidly. (Bybee, 2002:271)

The claim here is inherently one of change, rather than stable variation; the evidence brought forth in support of it, though, is not always clearly drawn from change in progress. For example, Bybee points to higher rates of coronal stop deletion in high-frequency word-final consonant clusters as support for frequency-conditioned lexical diffusion of sound change. There has never been any evidence, however, that coronal stop deletion is a change in progress in any North American dialect, with Guy calling even the notion that it might be a relic of an older change in progress “unfounded speculation” (1980:3). Given the assiduousness with which sociolinguists seek out change in progress and the high-profile status of coronal stop deletion in the sociolinguistic literature, it seems unlikely that such evidence

1 Thanks to Gene Buckley, Bill Labov, Joe Fruehwald, Constantine Lignos, Aaron Ecay, Georgia Zellou, and Charles Boberg for their comments on various parts of this project, as well as audiences at the 2013 LSA meeting and WCCFL 31.

has simply been overlooked. Even if coronal stop deletion were a change, albeit one proceeding at such a glacial pace that it could not be observed, a simple frequency effect observed in a snapshot of a change in progress is, as Pierrehumbert points out, “not enough in itself to argue for long-term storage of word-specific allophone detail” (Pierrehumbert, 2002:108). This is because frequency effects in production can in principle be implemented online through processing mechanisms such as spreading activation, without requiring such mechanisms to impact the phonological representation. The argument for exemplar theory, then, depends crucially on the gradual accumulation of usage-based phonetic differences between words over the course of a change. Note that unlike Bybee, who limits her discussion to reductive sound change (while suggesting that all sound change is ultimately reductive (2002:268)), Pierrehumbert explicitly extends the claim that frequent words lead sound change to any kind of gradient phonetic change, stating that “any systematic bias on the allophonic outcome would incrementally impact high frequency words at a greater rate than low frequency words” (2002:118). Just as frequent words that undergo reduction in speech should end up being more reduced in the phonetics inherent to their representation, frequent words that are undergoing non-reductive sound change (for example, the raising of /ey/ along the front diagonal in Philadelphia (Labov et al., 2013)) should accumulate advanced tokens more quickly than their less-frequent counterparts.

One area in which the search for word-specific phonetics has been pursued is homonyms. Vowel pronunciation in particular is highly sensitive to the conditioning effect of the surrounding phonological environment, meaning that comparing non-homonyms often leaves open the possibility that any observed differences are merely due to subtle coarticulation in production. Homonyms are therefore a key test for the existence of word-specific phonetics because the phonological environment is controlled. An early use of homonym differences to argue for lexical effects in sound change comes from Cheng & Wang (1977), who give twelve examples of homonym pairs that split into different phonemic categories in the development of Middle Chinese tone III in the modern Chao Zhou dialect. Labov does not find the same type of split across the homonym pairs know/no and twoltoo using data from Philadelphia (1994) and the Atlas of North American English (2010). On the other hand, Lavoie shows that four/for are reduced differently in natural speech. Johnson (2007) demonstrates that for 18 homonym pairs or sets the most common pronunciation variants are different, and Gahl (2008) finds length differences between frequent and infrequent members of homonym pairs.

A recent paper by Drager (2011) serves as the inspiration for the current study in its use of the homonym set that I will refer to as LIKE. The word like can be a lexical verb, a discourse marker, a quotative, or one of several other grammatical parts of speech, which I will discuss in section 2. Drager, focusing on the three-way distinction between the discourse marker, quotative, and grammatical (including verbal) functions of LIKE, demonstrates that in New Zealand English, “some of the lemma-based phonetic variation is socially conditioned and some of it is linked to the speaker-specific probability of producing the word” (2011:704) (with speaker-specific probability being one measure of frequency). By showing that the elements of the LIKE homonym set can be phonetically differentiated (in this case, by the consonantal elements of /l/-length and /k/-release), Drager sets us up to ask whether there is ever a case where they are not. I argue in this paper that there is such a case: the raising of the nucleus of /ay/ in Philadelphia. The raising of /ay/ before voiceless consonants is a feature found in a number of North American dialects, including Canada and the Inland North. In Philadelphia, the increasing phonetic differentiation of the /ay/ nucleus in voiceless environments from the stable low position of the /ay/ nucleus before voiced consonants and word-finally is a regular sound change that began in the early decades of the 20th century (Labov, 2001; Labov et al., 2013). The vowel in LIKE, of course, undergoes /ay/-raising due to the voicelessness of the /k/, meaning that Drager’s clever juxtaposition of the various functions of LIKE can be exploited in data from Philadelphia to evaluate word-specific effects on sound change in progress. The following sections will be dedicated to showing that, despite order-of-magnitude advantages in frequency, the most frequent LIKE homonyms do not take the lead in /ay/-raising.

This study is not the first to suggest that sometimes frequency effects fail to arise. In addition to the contributions from Labov, Dinkin (2008), Abramowicz (2007), and Walker (2012) all fail to find frequency effects for at least some of the variables they consider. But to my knowledge what has not

---

2 I adopt the phonological notation from Labov et al. (2006), in which /ay/ is the vowel in PRICE.
yet been pursued, likely because of limits on the available data sources, is a study of frequency and sound change across the entire course of a change. This is an important step, though, given that the claim of frequency conditioning sound change is a fundamentally diachronic one. Even the apparent time hypothesis, where older speakers are assumed to use a more conservative form of the language than younger ones, has not yet been used to demonstrate lexical divergence through the accumulation of phonetic differences during sound change. The development of the Philadelphia Neighborhood Corpus provides a sufficient time depth to observe the emergence of /ay/-raising from beginning to end. I contend that, if we take seriously Pierrehumbert’s caution about over-interpreting surface frequency effects, such a perspective is necessary to truly test the hypothesis of frequent words leading change.

At the same time as I broaden the time scale of the investigation, I also narrow the perspective lexically. The homonym studies described above notwithstanding, investigations into the role of frequency in sound change often consist of a proactive search to find the expected frequency effects. The net is set as wide as possible, with the target being the operation of a sound change over the entire lexicon. Here I take the opposite approach: I look at a very narrowly-defined context for the change, in which I fail to find the expected frequency effect. If we seek to falsify the strong Neogrammarian position on the regularity of sound change, we need to show that lexical divergence can arise. This task has already been thoroughly dispensed with; as Labov acknowledges, “we have arrived at a situation where no reasonable person can maintain what might be called the Neogrammarian dogma: that sound change is always gradual, always regular, affecting all words at the same time” (1981:271). On the other hand, if we seek to falsify the strong exemplar theoretic position, we instead need to show that it is possible for sound change to be insensitive to frequency in some context where the frequency effects would otherwise be expected to arise. This is because the prediction of exemplar theory is not merely that frequency can affect sound change, but rather that it should. If frequency effects fail to arise, even only sometimes, it is a challenge for a pure exemplar theoretic model.

2. Data and methods

2.1. The Philadelphia Neighborhood Corpus

The data for this study come from the Philadelphia Neighborhood Corpus of LING560 Studies (PNC) (Labov & Rosenfelder, 2011). The corpus comprises sociolinguistic interviews conducted by students at the University of Pennsylvania in a course called “LING560: Study of the Speech Community.” This course has been taught by William Labov, sometimes in conjunction with Gillian Sankoff and Anthony Kroch, since 1972. As the class continues to be taught biennially, fieldwork is ongoing; to date, 379 of the 1,107 recordings made in LING560 have been transcribed and included in the PNC. Labov et al. (2013) report that these transcribed interviews yield 889,000 vowel tokens when the transcripts are forced-aligned and the vowel formant measures extracted automatically using the Forced Alignment and Vowel Extraction (FAVE) suite (Rosenfelder et al., 2011). The study of distinct LIKE homonyms, however, still requires hand-coding, restricting how much data can be analyzed in a reasonable timeframe. Out of the pool of transcribed interviews I selected a sample of 37 white, upper working class speakers who were interviewed during the 1970s or the 2000s, with four age groups from each interview decade. Within each group the sexes of the speakers are balanced as closely as possible. In cases where there are fewer speakers of an age or sex category compared to the same category in the other interview decade, it is due to a lack of suitable speakers among the transcribed interviews. The number of speakers in each cell of the sample is given in table 1.

2.2. Coding LIKE

The coding of different functions LIKE was done by hand from the interview transcripts. Following D’Arcy’s (2005) extensive dissertation on the syntax of the discourse marker form of LIKE as well as the guidance of the Cambridge Grammar of the English Language (Huddleston & Pullum, 2002), I coded all instances of the wordform LIKE as belonging to one of the grammatical or discourse-pragmatic categories exemplified in table 2. Quotative LIKE had to be excluded because, although it is prevalent in more recent interviews, it was used only once in all 20 interviews conducted in the 1970s, making it impossible to trace its diachronic trajectory. Therefore, the types of LIKE that will be considered in
Table 1: Sample of PNC speakers included in this study

<table>
<thead>
<tr>
<th>Age at interview</th>
<th>Interviewed 1970s</th>
<th>Interviewed 2000s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Over 60</td>
<td>3 male / 3 female</td>
<td>3 male / 3 female</td>
</tr>
<tr>
<td>40-59</td>
<td>3 male / 3 female</td>
<td>2 male / 2 female</td>
</tr>
<tr>
<td>18-39</td>
<td>3 male / 3 female</td>
<td>3 male / 3 female</td>
</tr>
<tr>
<td>Under 18</td>
<td>1 male / 1 female</td>
<td>0 male / 1 female</td>
</tr>
</tbody>
</table>

Table 2: Functions of LIKE (examples from PNC subjects PH73-2-1, PH00-1-5, and PH06-2-3)

<table>
<thead>
<tr>
<th>Function</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical verb</td>
<td>I don’t LIKE the taste of beer anyways</td>
</tr>
<tr>
<td>Preposition</td>
<td>The tripe itself is almost bland as such, LIKE eating Jello.</td>
</tr>
<tr>
<td>Conjunction</td>
<td>But uh they didn’t go up together LIKE they used to.</td>
</tr>
<tr>
<td>Adjective</td>
<td>I don’t know if he’s exactly LIKE his father.</td>
</tr>
<tr>
<td>Quotative</td>
<td>And she’s LIKE, ”Let’s go to this fortune teller.”</td>
</tr>
<tr>
<td>Discourse marker</td>
<td>Um LIKE we used to play a lot of running games you know.</td>
</tr>
</tbody>
</table>

the results section are the lexical verb, the preposition, the conjunction, the adjective, and the discourse marker.

2.3. Vowel measurement

The transcribed interviews that make up the PNC have been forced-aligned using the FAVE-align program, a version of the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008) that is adapted for use with sociolinguistic interviews. The aligned tiers are then subjected to automated formant extraction using FAVE-align’s sister program, FAVE-extract. FAVE-extract uses Linear Predictive Coding in Praat to measure F1, F2, and F3 in Hertz (Hz) at a specified point in the trajectory of the vowel, then rechecks the measurements to exclude gross measurement errors (see Evanini (2009) and Labov et al. (2013) for details). Although the default setting for FAVE-extract excludes vowels shorter than 50 milliseconds in duration, I included all vowels to avoid the possible pitfall of obscuring a potential interaction between homonym frequency and duration. Formant measurements were normalized using Lobanov’s (1971) method, which is an intra-speaker z-score (see Adank (2003) for a comparison of normalization techniques). Since raising is primarily a change in the nucleus height of /ay/, I will take the normalized F1 value to be the most useful measure of /ay/-raising.

2.4. Quantifying frequency

A methodological challenge for any study of frequency effects in phonetics is that there is not yet a solid consensus on the most appropriate quantitative treatment of frequency itself. As Erker and Guy note, “it is not always clear how frequency is best defined—locally or globally, continuously or discretely, by lemma, form, or collocation, at what level of granularity, and so on” (Erker & Guy, 2012:527). A common approach, which Erker and Guy find support for in the conditioning of Spanish subject personal pronoun expression, is to bin lexical items into high- and low-frequency categories. This is the type of approach Bybee assumes when she argues that Labov’s homonym pair results (1994; 2010) are “not definitive since all of the words used occurred three or more times in the interview and thus must be considered high frequency” (2002:267). She does not, however, give justification for setting an arbitrary cut-off point at three tokens per interview. Furthermore, at least one often-cited result in support of a binary approach to frequency, from Alegre & Gordon (1999), has since been suggested to be the outcome of a statistical error (Lignos & Gorman, forthcoming). I assume here that in the absence of compelling evidence to justify a meaningful high/low cut-off, it is reasonable to expect that frequency differences as large as 129 versus 1149 occurrences should produce phonetic differentiation in sound
Table 3: Within-dataset frequency of the LIKE functions

<table>
<thead>
<tr>
<th>Function</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical verb</td>
<td>213</td>
</tr>
<tr>
<td>Preposition</td>
<td>274</td>
</tr>
<tr>
<td>Conjunction</td>
<td>129</td>
</tr>
<tr>
<td>Adjective</td>
<td>138</td>
</tr>
<tr>
<td>Discourse marker</td>
<td>1149</td>
</tr>
</tbody>
</table>

change by the exemplar theoretic mechanisms Bybee (2002) and Pierrehumbert (2002) describe. On the topic of local versus global frequency measures, my methodological decisions are constrained by the nature of the case study. In general it would be preferable to use global frequency norms taken from large corpora because they are more stable. None of the frequency norms available, though, differentiate between the LIKE homonyms that are the topic of this paper. Consequently I will rely on the number of occurrences of each homonym within the dataset used here. These counts are presented in table 3. Although they are surely not ideal, I depend on them only to give a general picture of the relative frequencies of the homonyms.

3. Results on /ay/ raising

Labov et al. (2013) show that the PNC captures essentially the full span of the change from a low to a centralized /ay/ nucleus before voiceless segments. As a sanity check, I begin by confirming that my sample is generally representative of Philadelphia in that the speakers participate in the diachronic process of /ay/-raising. I code all instances of the vowel /ay/, excluding function words and forms of LIKE, as being in the phonological environment of a voiced or voiceless following segment. The normalized F1 values in the two environments over time are shown in figure 1.

![Figure 1](image1.png)

Figure 1: /ay/-raising before voiceless segments (N=1499) and stability before voiced segments (N=2823) by year of birth. LOESS fit.

Next I turn to the crucial question of whether the different functions of LIKE show different degrees or rates of raising. Figure 2 shows the F1 values for each of the functions of LIKE over time.
We see in figure 2 that the adjective, conjunction, discourse-marker, and preposition functions of LIKE cluster so closely as to be nearly indistinguishable, while the lexical verb function of LIKE apparently lags behind. In figure 2 I focus on the oldest age group (speakers over the age of 60 when they were interviewed in the 1970s) and youngest age group (speakers under the age of 40 when they were interviewed in the 2000s) to isolate the distinct behavior of the lexical verb at either endpoint of the change. The small apparent lexical effect that exists before the change gets underway, rather than being exaggerated by the progression of the change, is gone by the time it enters its later stages.

Figure 3: /ay/-raising for different functions of LIKE: speakers over 60 interviewed in 1970s (N=144) versus speakers under 40 interviewed in 2000s (N=542).
3.1. The role of function word reduction

Treating each homonym of LIKE as distinct overlooks a relevant generalization: the adjective, conjunction, discourse marker, and preposition forms of LIKE have in common that they are all function words, while the lexical verb LIKE is the only content word. Notice that this distinction aligns with the difference between the lower and higher /ay/ vowels early in the change. I suggest that the function word status of the more raised LIKE forms early in the change is sufficient to account for that slight amount of raising relative to the lexical verb. Lexical verb LIKE has an F1 value of around 2 at the beginning of the 20th century, which is the same value as we see for the allophone in the non-raising environment throughout the change. At the beginning of the change, then, this is the target for the raised allophone as well. Lexical verb LIKE, being a content word that bears lexical stress, generally displays faithful achievement of this target. In contrast, I suggest that the reason the other forms of LIKE are subject to some degree of centralization is that they are function words. Function words are typically shorter, unstressed, and phonetically reduced (Bergem, 1993; Jurafsky et al., 1998; Bell et al., 2009). The gap between verbal LIKE and the other LIKES early in the change can thus be accounted for simply as a reflection of function word reduction in production. By the end of the change, the F1 target of raised /ay/ (as illustrated by the quality of the verb) is near 0, meaning almost entirely centralized. There is therefore no longer room for function word reduction—in other words, this is a ceiling effect. This accounts for the loss of the difference late in the change.

Further support for this account can be found in the behavior of the other allophone of the /ay/ diphthong. As seen above in figure 1, /ay/ before voiced consonants does not undergo raising in Philadelphia. Using data from the 20 speakers interviewed in the 1970s to reflect the earlier part of the change, I divide the /ay/ measurements from voiced contexts into content and function words. Figure 4 indicates that the difference between the content and function words for /ay/ in the stable voiced environment is in the same direction and of a similar magnitude as the difference between the content and function words for /ay/ in the changing voiceless environment. In both cases the median of the function word nuclei is slightly higher than those of the content words, but still within the interquartile range. This slight raising of /ay/ before voiced segments in function words cannot be attributed to change, as this allophone is stable. Indeed, the failure of function word reduction in the voiced context to snowball into a change comparable to that in the voiceless context is itself a puzzle for exemplar theory. As Pierrehumbert asks regarding very small phonetic differences, “So why don’t they pile up?” (2006:523).

Figure 4: /ay/ raising in content and function words for LIKE (N=873) and voiced /ay/ (N=5926), from 1970s interviews.
4. Discussion

The exemplar theoretic model of phonology, at least in its strong version, predicts that phonetic change should accrue the most quickly in the most frequent words. The results in the previous section are difficult to reconcile, though, with the claim that frequent words lead sound change. The adjective, conjunction, discourse marker, and preposition forms of LIKE are in lockstep throughout the entire course of the change, despite order-of-magnitude differences in their within-dataset frequencies. Meanwhile, lexical verb LIKE, which falls in the middle of the frequency range, is lower than the other homonyms at the beginning of the change but catches up by the end. This lexical effect, the only one present in the data at hand, bears no relation to frequency but rather corresponds with the grammatical function of the lexical items. It therefore seems to be the type of effect that Pierrehumbert (2002) points out can be handled online in a non-exemplar-based framework, being comparable to a similar effect in the stable /ay/ allophone and plausibly mediated by function word prosody. Setting aside frequency, the rate of change over time in the lexical verb also operates in the opposite direction of the exemplar theoretic prediction—the lexical items that occur most frequently in an unstressed and therefore reduced context actually undergo the raising change more slowly from beginning to end, while the change in the lexical verb has a steeper slope despite starting out at a lower point.

The problem that these results pose for exemplar theory is that it provides no native mechanism to constrain word-by-word phonetic divergence. It is, of course, a difficult problem for all theories of sound change that sounds sometimes (often) stay the same. But while some approaches have the problem of not being able to predict when sound change will and will not occur, exemplar theory has the deeper problem that the perpetually-emergent nature of its lexicon makes it difficult to account for any stability at all. One direction we might go in reconciling the findings that have been marshaled in support of exemplar theory with the finding here that frequency need not cause phonetic divergence is towards what Pierrehumbert (2002; 2006) terms hybrid models. In such models “a phonological coding level intervenes between the lexicon and the parametric phonetic description” (Pierrehumbert, 2006:523). Although such models likely have the new problem of predicting when differentiation in language use experience will accrue to the phonetic description and when it will “bounce off” the phonological level, this is parallel to the traditional problem of explaining the occurrence and nonoccurrence of sound change. The parallelism is sufficiently strong that one might ask whether hybrid exemplar models are qualitatively unique in their description of the human linguistic capacity or whether they are simply one of the more probability-driven approaches to the phonetic implementation of abstract, and phonetically-impermeable, phonological categories. I suggest that further research on the full trajectory of sound changes, enabled by the increasing availability of large-scale sociolinguistic corpora from specific speech communities, will play a key role in answering such questions and shedding light on the nature of the phonology and the lexicon.

References


