Entry title: Lexical frequency and diffusion

First author:
Vsevolod Kapatsinski, University of Oregon, vkapatsi@uoregon.edu

Abstract

Linguistic changes often diffuse through the lexicon, spreading from word to word. The direction of diffusion depends on word frequency and the mechanisms driving the change. This chapter critically reviews research on frequency effects in language change from a usage-based perspective. Reductive sound changes are argued to affect frequent words first. In contrast to much current literature, it is argued that the frequency effect need not increase when the change accelerates for this explanation to hold. Frequent words are also argued to be more likely to be extended to novel uses, because of their higher accessibility. This accounts for certain kinds of semantic change (e.g., the changes we see in grammaticalization) and morphological paradigm leveling. Analogical changes and those due to misperception are argued to preferentially target low-frequency words. An important complication arises because words occur in contexts that either favor or disfavor a change. Words that frequently occur in change-favoring contexts and infrequently occur in change-disfavoring contexts are the words that are most likely to change and to change fastest. The direction of change in semantics depends almost entirely on the contexts in which the word occurs: words that frequently occur in a context that licenses a certain pragmatic inference can become associated with that inference, carrying it to other contexts. That is, social and linguistic contexts host bits of form and meaning that can stick to words (and other expressions) that frequent those contexts. Current methodological issues with studying the role of frequency in lexical diffusion are summarized.

Keywords

[lexical diffusion, word frequency, sound change, semantic change, semantic extension, paradigm leveling, frequency in favorable contexts, reduction, computational models of change]

Main text

[A] Introduction

Lexical diffusion refers to the spread of a change through the lexicon. Although the term itself was coined by Wang (1969), debates on the role of word frequency in language change date back to the late 19th century. At that time, the debate centered on the newly proposed Neogrammarian Hypothesis of regular *sound* change. According to the Neogrammarian Hypothesis, change is localized to sounds and not words, and all words containing a sound in a certain phonological environment change together (Osthoff and Brugmann 1878). This hypothesis was subsequently reaffirmed by Bloomfield (1933) and, to a large extent, Labov (1981, 2020).

From the beginning, the Neogrammarian hypothesis was opposed by the dialectological tradition in which each word had its own history and change diffuses gradually through the lexicon. The role of word frequency was emphasized in this tradition by Schuchardt's (1885) famous statement that "rarely used words drag behind; very frequently used ones hurry ahead" in sound change. The idea of lexical diffusion was revived by Wang (1969), and the effects of frequency began to be investigated by Fidelholtz (1975) and Hooper/Bybee (1976). As large longitudinal speech corpora became available in the last decade, and linguistics truly embraced inferential statistics, these studies have grown dramatically in sophistication (e.g., Cohen Priva and Gleason 2020; Forrest 2017; Foulkes and Hay 2016; Hay et al. 2015; Labov 2020; Sóskuthy and Hay 2017). I will argue that these studies have upheld the idea of sound change being led by frequent words when the change is reductive, i.e., when it is driven by a pressure to optimize articulation.

Whereas Schuchardt (1885) emphasized change starting in frequent words, Paul (1877), Wheeler (1887), and Kruszewski (1890), proposed that some changes diffuse in the opposite direction, starting in low-frequency words and affecting high-frequency words last (if at all). These are analogical changes, in which words with weak memory representations succumb to peer pressure from their more numerous neighbors that behave differently. The effect of frequency on analogical change has been less controversial, because it is difficult to imagine a view of language acquisition in which it would not arise.

This chapter review the evidence for these effects and addresses some complications that arise when we take into account the fact that change is not driven by reduction or imperfect learning alone.

[B] Context matters

Because analogical changes affect rare words first, and reductive changes affect frequent words first, the primary motivation for a change can often be ascertained from the direction in which a change is diffusing through the lexicon (Bybee 2001). To a large extent, this is true. However, some complications arise because production choices are multiply determined.

First, pronunciation variants -- once introduced -- acquire social meanings and are then subject to social selection (e.g., Labov 1981). The role of word frequency in social selection is poorly understood. On the one hand, it has been shown that speakers find it more difficult to imitate the pronunciation of another speaker when they are producing a frequent or predictable word (Goldinger 1998; Manker 2019). This would suggest that social selection should affect frequent words less than infrequent words because -- even when intending to imitate a role model -- a speaker finds it harder to shift their entrenched production habits for a frequent word. On the other hand, innovative pronunciations of (locally) frequent words can become salient social stereotypes (e.g., Hay, Nolan and Drager 2006). For example, Oregonians insist on a schwa in the last syllable of Oregon, and not reducing the vowel is a salient marker of being an out-of-stater. Reducing a toponym is an especially good cue to local identity because it suggests that the toponym has been frequent in the speaker's experience. Because there are many ways to reduce a word, reducing it 'properly' also requires experience with the local speech community. Since speakers have more experience with a frequent word, allowing them to learn such distinct pronunciations and their connotations, speakers are likely to develop a wider repertoire of distinct ways to pronounce a frequent word, to be deployed in different social situations (Kapatsinski, Easterday and Bybee 2020). Some evidence for frequent words showing a wider range of intentional variability in pronunciation is provided by the fact that some frequent words undergo special reduction unattested in other words and this special reduction happens only in certain distinct uses of a word (Barth 2015; Bybee & Scheibman 1999; Bybee, File-Muriel and Napoleão de Souza 2016). For example, *I don't know* can be reduced to a nasal schwa when it functions as a marker of uncertainty. In addition, the durations of frequent words appear to be more strongly affected by discourse structure, lengthening more than rare words when they are mentioned for the first time within a discourse episode (Bell et al. 2009; Vajrabhaya 2015).

Second, a crucial complication is that changes are favored and disfavored by various contexts -- social as well as language-internal (phonological and grammatical). That is, some tokens of a word are more likely to contain an innovative pronunciation. For example, the deletion of a /t/ at the end of a word is more likely before a consonant than before a vowel, and is more likely in some dialects than in others (Coetzee and Pater 2011). At the same time, changes also accrete to words (Alba 2005; Brown 2004; Bybee 2002; Pierrehumbert 2001, 2002). That is, words that frequently occur in contexts that favor an innovative pronunciation are also more likely than other words to be pronounced that way in other contexts (Alba 2005; Brown 2004; E. K. Brown 2020; E. K. Brown and Alba 2017; Brown and Raymond 2012; Brown and Shin 2022; Bybee 2002, 2017; Forrest 2017; Raymond and Brown 2012; Raymond, Brown and Healy 2016, Tang and Shaw 2021 though cf. Cohen-Goldberg 2015 for a null result). For example, /t/-final words that tend to occur before consonants are also more likely to be pronounced without a /t/ before vowels (Bybee 2002). In semantics, words that frequently occur in a context that licenses a certain pragmatic inference become associated with that inference even when they occur outside of the licensing context (Sinclair 1991; Traugott and Dasher 2002; Traugott and König 1991). For example, the word *cause* has tended to occur in the context of causing something bad, and has developed a semantic prosody such that it sounds odd to *cause happiness* and *causes of language learning* is worse than *causes of language impairment*. The French word *pas* used to mean 'step' but has come to mean 'not' from frequent occurrence in the phrase *ne pas* 'not a step' in which it was the more acoustically salient element. In syntax, verbs that often occur in pragmatic contexts that favor overt pronoun expression in Spanish come to favor overt pronoun expression in other contexts (Orozco and Hurtado 2020; Travis and Torres Cacollos 2021).

These effects suggest that a word's propensity to undergo a change is affected not by overall word frequency but rather by a word's frequency of occurrence *in contexts that favor the change*. Frequent occurrence in (some) other contexts may in fact disfavor the change. Accordingly, a large number of studies have found that change correlates with what Raymond and Brown (2012) call *frequency in favorable contexts* (FFC), which is actually a difference between frequency in the contexts that favor a change and frequency in contexts that disfavor it (Alba 2005; Brown 2004, 2013; E. K. Brown and Alba 2017; Brown and Raymond 2012; Forrest 2017; Raymond and Brown 2012; Raymond, Brown and Healy 2016; Sóskuthy and Hay 2017). For example, Sóskuthy and Hay (2017) examined a corpus of speech recordings of New Zealand English for speakers with birth years spanning 130 years, and found that words that frequently occurred utterance-finally -- a context that causes final lengthening as articulators slow down to a stop -- also lengthened in utterance-medial contexts.

Because frequent words also tend to be the words that frequently occur in change-favoring contexts, frequency is often a good proxy for frequency in change-favoring contexts. For example, reductive changes are favored by casual speech and frequent words are the ones more likely to be used in casual speech. However, some of the studies on FFC referenced above have tried to differentiate FFC and overall frequency by putting both in a regression model. Some of these studies found an effect of FFC in the absence of an overall frequency effect (e.g., Raymond, Brown and Healy 2016, but cf. Forrest 2017). However, the inherent collinearity between FFC and frequency makes interpretation of these conflicting results a little tricky. Frequency is usually log transformed for analysis purposes, as log frequency yields stronger effects than raw frequency, either because practice has diminishing returns or simply because raw frequency distributions are highly skewed. On the log scale, FFC is $\log(Freq_{favorable})–\log(Freq_{Unfavorable})$, and overall frequency is $\log(Freq_{favorable})+\log(Freq_{Unfavorable})$. Therefore, a regression model with FFC and overall frequency in the model is $FFC + Freq = \log(Freq_{favorable}) + \log(Freq_{Unfavorable}) + \log(Freq_{favorable}) + \log(Freq_{Unfavorable}) = 2 * \log(Freq_{favorable})$. Consequently, an effect of FFC in the absence of an effect of overall frequency suggests that frequency in unfavorable contexts slows down the change, whereas an effect of FFC alongside an effect of overall frequency suggests that frequency in unfavorable contexts simply does not matter much.

This variability in results is therefore likely due to some unfavorable contexts being more unfavorable than others. For example, if /t/ deletion is socially stigmatized in a speech community, words that frequently occur in formal contexts would be less reduced than those that occur in such contexts infrequently. In contrast, when /t/ deletion is not stigmatized, additional occurrences in formal contexts may favor rather than disfavor reduction by providing additional opportunities for production practice. In other words, frequency in unfavorable contexts is likely to matter particularly when unfavorable contexts are associated with a preference for the conservative variant.

With these caveats in place, we now turn to a review of correlations between word frequency and susceptibility to change, and the mechanisms behind these correlations.

[C] Changes that tend to affect frequent words

Changes that affect frequent words first are particularly interesting because they are unlikely to be due to imperfect learning, or to the word losing its peculiarities under the influence of its phonological or semantic neighbours: these are most practiced, most entrenched members of the lexicon (Bybee 1985, 2001). And yet, some changes affect these words first. Two types of change are favored by high token frequency of a word:

1) changes that are attributable to automatization of articulation, and therefore mirror the sorts of reductions we see in casual or rapid speech (e.g., Baranowski and Turton 2020; Bybee, File-Muriel and Napoleão de Souza 2016; Bybee and Napoleão de Souza 2019; Dinkin 2008; Fidelholtz 1975; File-Muriel 2007, 2010; Forrest 2017; Gahl and Baayen 2022; Hay and Foulkes 2016; Hooper 1976; Lin, Beddor and Coetzee 2014; Sóskuthy and Hay 2017; Yun 2006; but cf. Labov 2020, Walker 2012 for null results), and

2) changes that are due to the high accessibility of a wordform that doesn't quite match the intended meaning in the moment of production, which involve the range of the wordform's uses expanding at the expense of less accessible alternatives (e.g., Bybee 2003; Bybee & Brewer 1980; De Smet 2016; Harmon and Kapatsinski 2017; Kapatsinski 2009; Koranda, Zettersten and MacDonald 2022; Tiersma 1982; Zipf 1949). We take these two types of changes in turn.

The usage-based approach to linguistics is particularly suited to accounting for these types of change because, by assuming that language structure emerges from language use, it can explain how the dynamics of lexical selection and learning within the production system influence change in the grammar (e.g., Bybee 2001, 2006; Lindblom et al. 1995; Kapatsinski 2018; Mowrey & Pagliuca 1995).

[C.1] Articulatorily-motivated sound change

We begin with sound change. There are multiple types of sound change that have different primary motivations and therefore spread across the lexicon differently (e.g., Bybee 2001, 2002; Hooper 1976; Phillips 1984, 2006). In this section, we are concerned with *articulatorily-motivated* sound change. These are reductive changes that appear to be motivated by optimization of a word's articulation with increasing practice. We can tell that a change is reductive because it mirrors the kinds of online reductions we see in rapid and casual speech and in situations where the word's meaning is highly predictable from context. Reduced duration has been persuasively argued to be a necessary criterion (Mowrey and Pagliuca 1995), and there is increasing evidence that a change in duration mediates other reductive changes (Cohen Priva and Gleason 2020). That is, other things about the pronunciation of a word might change as it is produced more quickly. These reductive changes can optimize articulation by:

1) increasing coarticulation / gestural overlap, especially of the anticipatory kind, because practice producing a sequence of articulations to express a meaning allows the speaker to increasingly anticipate upcoming articulations (Cohen Goldberg 2015; Mowrey and Pagliuca 1995);

2) reducing articulatory gesture magnitudes (Browman and Goldstein 1991; Mowrey and Pagliuca 1995), which sometimes culminates in the complete elimination of a gesture, and

3) smoothing out articulator trajectories and articulator velocity over time, so that articulators move at a relatively constant speed rather than in fits and starts (Kapatsinski 2018, Chapter 9; Sosnik et al. 2004).

The same change can often be seen as increased anticipatory coarticulation, reduction in gesture magnitude and smoothing out of the velocity profile. For example, the American English 'flapping' in *butter* (which often results in an approximant rather than a flap; Patterson and Connine 2001) reduces the magnitude of the alveolar closure gesture, but also increases coarticulation of the vowel-flap-vowel sequence. The influence of coarticulation can be seen even more clearly in the usual presence of voicing throughout the flap, even one that used to be a voiceless [t].

The finding that experience leads to increasing smoothness of the velocity profile of a moving articulator (Sosnik et al. 2004) can bring additional changes under the purview of reductive, articulatorily-motivated change by demonstrating that allowing a heavy articulator to travel a bit farther under the force of inertia can be easier than forcing it to abruptly change direction. For example, Tomaschek et al. (2013) show that speakers produce more extreme tongue body lowering in producing vowels in frequent words, where they need to be produced more quickly. Traveling farther can also optimize articulation when stopping earlier requires precise control. For example, stopping interdental fricatives in words like *this* and *that* (i.e., turning these into *dis* and *dat*) replaces an articulation that requires precise positioning of the tongue (close enough but not too close to generate turbulent airflow) by an articulation that can be produced by an uncontrolled ballistic movement (Kapatsinski 2018, Chapter 9).

Because our understanding of articulation and articulatory control is still evolving, it is often challenging to determine that a change is not reductive based entirely on its substance. It is therefore safest to assume that the kinds of changes we see across languages when predictable words are produced rapidly and casually are reductive changes.

Fidelholtz (1975) and Hooper (1976) conducted some of the first empirical studies testing Schuchardt's hypothesis that frequent words are ahead of infrequent words in sound change, finding support for this idea. For example, Hooper (1976) suggested that the second vowel in the rare word *mammary* is less likely to be reduced to a schwa than the second vowel in the moderately frequent *memory*, and pointed out the further reduction of the schwa to zero in the very frequent *every*. Although the methods used in these early studies would not satisfy the current standards in the field (Johnson 2009; Walker 2012), the facilitating role of frequency in such *reductive* changes has been held up by the body of work that followed (e.g., Baranowski and Turton 2020; Bybee, File-Muriel and Napoleão de Souza 2016; Bybee and Napoleão de Souza 2019; Forrest 2017; Gahl and Baayen 2022; Hay and Foulkes 2016; Lin, Beddor and Coetzee 2014; Sóskuthy and Hay 2017; Sóskuthy et al. 2015).

The simplest reason for why frequent words are more likely to be more reduced than infrequent words is because they are more likely to occur in reduction-favoring contexts than infrequent words. After all, frequency is average predictability (Gregory et al. 1999; Jurafsky et al. 2001), thus frequent words do not need to be articulated as carefully to be understood and are produced more quickly and casually. Frequent words are also more likely to be produced in informal situations: one does not deploy learned vocabulary in a bar fight. Therefore, if one analyzes the rate at which different words are reduced in a corpus, frequent words will be reduced at a higher rate than infrequent words.

To make this idea concrete, let us implement it in the following simple model of sound change. Imagine that there are two pronunciation variants of a sound, one of which is more reduced / easier to pronounce. Each generation of speakers learns the relative probabilities of the variants in the speech of the previous generation. Each generation also uses these variants under various contextual production pressures, which remain constant across generations and produce a positive relationship between frequency and reduction. This model is an oversimplification of Kapatsinski (2021) in that the learners do not infer a random effect of word identity when learning the language and is intended to be the simplest possible model that produces frequency-driven reduction.

Formally, for each generation, $p(\text{reduction})_{\text{Word}} = \text{logit}^{-1}(b_0 + b_{\text{Freq}}{*}\text{Freq}_{\text{Word}} + N(0,\text{sd}_{\text{Noise}}))$, where $p(\text{reduction})_{\text{Word}}$ is the probability of using the reduced variant for a certain word, b_0 is the probability of the reduced variant in the speech of the previous generation (as inferred by the next generation), b_{Freq} is the effect of various contextual pressures that make it more likely for speakers to choose the reduced variant for a frequent word compared to an infrequent word, and *N* is random (normally-distributed) 'noise', which represents all other influences on reduction. I put 'noise' in quotes because many of these influences are surely not random. However, as emphasized in McElreath (2020), the effect of many small influences summed together is normally-distributed. Freq is the (log-scaled) frequency of the word. logit^{-1} converts the sum from the log odds scale to the probability scale. The summing is done on the log odds scale because summing probabilities causes them to go outside the 0-1 interval rendering the result uninterpretable as a probability (Cedergren and Sankoff 1974). Word frequencies were the frequencies of verbs from the English Lexicon Project (Balota et al. 2007), as estimated from SUBTLEX. b_0 was set to –5 at the beginning, representing a low initial probability of the reduced variant; $\text{logit}^{-1}(-5) = .007$. b_{Freq} was set to 0.5. and sd_{Noise} to 0.1. The qualitative results do not depend on these parameters. Code is available at https://osf.io/kqhg3.

This simple model is sufficient to produce sound change led by frequent words: as shown in Figure 1, frequent words are more reduced than infrequent words for every generation, and the sound change progresses across generations, so that eventually (light dots) the reduced variant dominates.

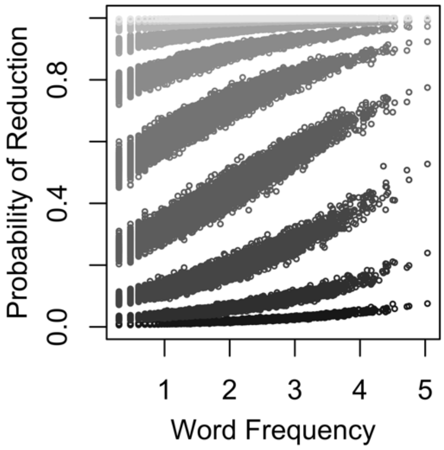

Figure 1. The relationship between word frequency and reduction across 10 generations (darkest = earliest) under the assumption that each generation infers an overall probability of reduction and then produces the variants based on that probability, frequency of the word (which has the same effect in all generations), and a constant random noise representing all other factors influencing reduction.

The change in Figure 1 appears to pick up speed: starting from the bottom, Generations 1 and 2 at the bottom look more alike than Generations 2 and 3, and then decelerate: Generations 8 and 9 at the top look more alike than Generations 2 and 3. That is, the change looks S-shaped. In tandem with this general dynamic, the effect of frequency appears to increase and then decrease: in Figure 1, the slope of the frequency effect in Generations 1 and 9 appears shallower than in Generation 4 (counting from the bottom). However, the effect of frequency was programmed to be the same for every generation on the log odds scale and indeed remains so (Figure 2).

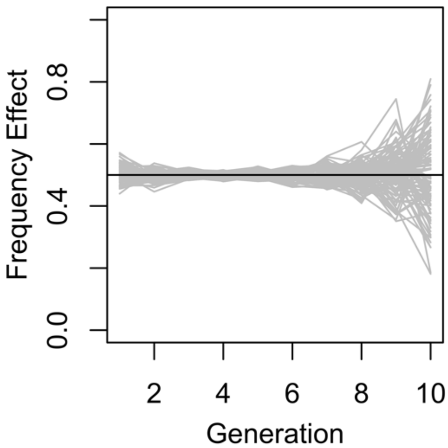

Figure 2. The effect of frequency by Generation. 100 grey lines are drawn for 100 replications of running the sound change through the 10 generations. Solid black line is drawn at the pre-programmed value of the frequency effect and shows the overall (lack of) trend. The increased variability at the edges is due to difficulties of inferring the overall reduction probability when it approaches 0 and 1 because 0 and 1 correspond to minus and plus infinity on the log odds scale.

The constant magnitude of the frequency effect in Figure 2 is important because it has been assumed that a reductive sound change led by frequent words means that the effect of word frequency on reduction should accelerate as a change picks up steam, and that the frequency effect remaining constant constitutes evidence against this hypothesis (Hay et al. 2015; Hay and Foulkes 2016; Labov 2020; Pierrehumbert 2001; Todd, Hay and Pierrehumbert 2019; Zellou and Tamminga 2014). For example, Zellou and Tamminga (2014) observed a decrease in nasal closure in words like *can* accompanied by an increase in nasalization of the preceding vowel in Philadelphia English. This would appear to be a reductive change due to increased anticipatory coarticulation, and an excellent confirmation of the prediction by Mowrey and Pagliuca (1995) that anticipated gestures will move forward in time. As would be expected, Zellou and Tamminga also observed more vowel nasalization in frequent words than in infrequent words. However, the size of the frequency effect did not change across generations, leading them to conclude that "synchronic frequency effects result from cognitive mechanisms independent from the triggers and drivers of community-level change".

However, the change in Figures 1 and 2 is due to the synchronic frequency effect, even though the effect of frequency is constant across generations: if there were no frequency effect in online production pushing speakers to choose the reduced variant, the change would not happen (Figure 3, right). Each generation would faithfully infer the (low) probability of the reduced variant, and reproduce it in their own speech. It is because of the online effect of frequency on reduction that the overall reduction probability is incremented by each generation, allowing the change to progress (Figure 1). Learning of the overall probability of reduction, which represents the community norm of the previous generation, is also necessary for incrementation. Without it, speakers would always reduce frequent words more than infrequent words but the change would not progress, resulting in stable variation (Figure 3, left). The two influences independently affect production choices, but reductive sound change requires both influences to be there.

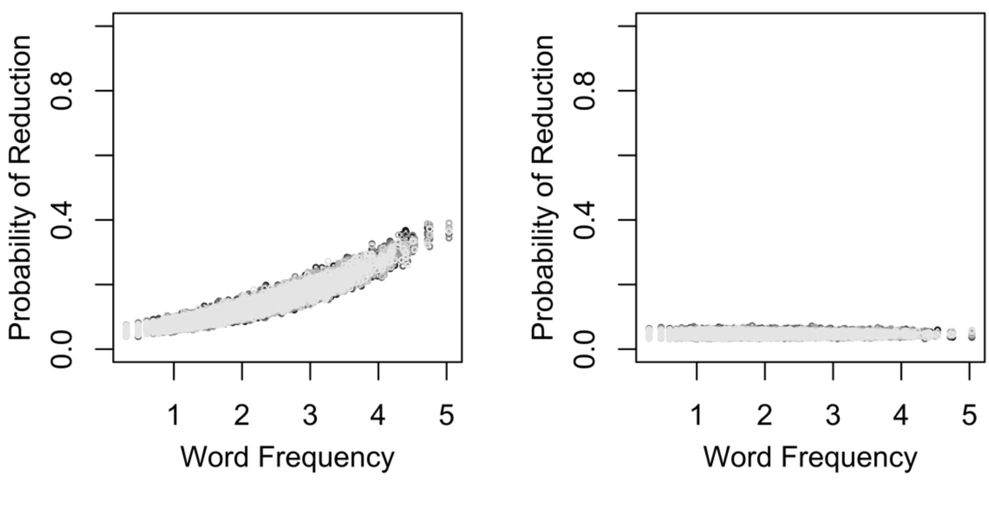

Figure 3. No reductive change happens without speakers learning to reduce at the community rate (left) or the synchronic frequency effect on reduction (right).

The belief that the effect of frequency on reduction should increase as a sound change accelerates seems to be due to previous simulations of articulatorily-driven sound change (e.g., Labov 2020; Pierrehumbert 2001; Todd, Hay and Pierrehumbert 2019) being initialized with all words behaving alike. This initialization can be rhetorically effective to show that a frequency effect would emerge from processing pressures even if all words were initially the same. However, it is an unrealistic assumption for reductive sound changes: the online production pressures are always there, active in every instance of language use, and there is no initial state of the language in which frequent words are expected not to be more reduced than infrequent words. Rather, frequent words should always be ahead of rare words as they drift across the space of possible pronunciations (e.g., Bybee 2001; Mowrey and Pagliuca 1995). In particular, we would not expect to see a language in which frequent words are pronounced as slowly as rare words, and to the extent that rapid pronunciation is what causes other sorts of reduction (Lindblom 1963; Cohen Priva and Gleason 2020), we would always expect more online reduction in frequent words.

Based on this hypothesis of omnipresence of online reduction of frequent words, the first generation in the simulation above already showed an online reductive effect of frequency. I believe this to be a more faithful reflection of the theory that sound change starting in frequent words is reductive: the production pressure to reduce frequent words is always there, but at some point the reduced pronunciation becomes socially marked and speakers begin to learn when, and how often to use that pronunciation, which results in change. As a result, the overall pressure in favor of the reduced variant *does* increase as the change progresses, but only because it is increasingly socially favored. This social pressure need not change the influence of experience on articulatory optimization for the sound change to progress, or to be driven by frequent words.

Note that the frequency effect in Figure 3 (left) and the frequency effect in Figure 1 are the same exact frequency effect. The only difference between a change in progress (Figure 1) and stable variation with online reduction (Figure 3, left) is whether the overall reduction probability is part of the community norm. That is, change happens when an innovative pronunciation becomes something you learn to do because others in your community do it. In other words, when change is happening, the choice of pronunciation is under partial control of the grammar. In Figure 3 (left), it is not, so new speakers of the language do not learn to reduce at the community rate while in Figure 1 it is, and they do. From this perspective, studying frequency effects in situations of stable variation (e.g., Forrest 2017) is just as informative about the frequency effects driving reductive sound change as studying them in situations of change (cf. Labov 2020; Todd, Hay and Pierrehumbert 2019 for recent statements to the contrary). As expected from this perspective, studies of stable variation reveal the same experience-based factors being responsible for the variation as studies of reductive change (e.g., Bybee 2002; Bybee, File-Muriel and Napoleão de Souza 2016; Forrest 2017).

If the reductive effects of frequency are equally strong in stable variation and language change, why do frequency effects sometimes increase in magnitude as a change speeds up (e.g., Hay and Foulkes 2016)? In the simple model above, the social cache of the innovative variant increases as it becomes more prevalent, which causes speakers to increasingly favor it. As mentioned earlier, it is possible that frequent words are more subject to social selection because they often serve as social shibboleths or stereotypes (Hay, Nolan and Drager 2006; Drager and Kirtley 2016; Grondelaers and Marzo 2023). If so, the innovative variant should be increasingly favored by frequent words as it increases in prevalence.

From this perspective, the effects of frequency on a reductive change are also expected to decrease when the change decelerates for two reasons. The slowdown is likely to happen either 1) because the innovative pronunciation has acquired social stigma, which can cause the change to slow down or even reverse (e.g., Delforge 2009 shows reversal of unstressed vowel devoicing as a result of social stigma in Quito Spanish; Dodsworth and Kohn 2012 and Labov, Rosenfelder and Fruehwald 2013 show reversal of vowel shifts in some varieties of US English), or 2) because articulatory efficiency approaches a local maximum as reduction proceeds -- if reductive changes improve articulatory efficiency, it is logical to assume that reductive pressure should weaken as a result of reductive change. If the slowdown of change is due to the innovation acquiring social stigma, frequency effects are expected to reduce or even reverse for the same reason they increase when the change accelerates -- because frequent words are more subject to social selection. If the change slows down because production efficiency approaches a local maximum, and the reductive pressure weakens, frequency effects should decrease because they arise from this reductive pressure (see Kapatsinski 2021 for modeling results showing the frequency effect reducing when a change approaches a stable state).

To the extent that the effect of word frequency is due to online reduction, and online reduction is based on the word's predictability (Jurafsky et al. 2001), frequency is merely a proxy for a word's average predictability in context. For example, the word *legomena* has very low frequency but is almost always predictable from its preceding context, since it almost always occurs after the word *hapax*. Thus, several researchers have recently investigated various measures of predictability and *informativity*, which averages predictability across the contexts in which a word(form) occurs (e.g., Alba 2008; Barth 2019; Cohen Goldberg 2015; Cohen Priva 2017a, 2017b; Hashimoto 2021; Kurumada and Grimm 2019; Purse, Fruehwald and Tamminga 2022; Seyfarth 2014; Shaw and Kawahara 2019; Sóskuthy and Hay 2017; Tang and Shaw 2021; Turnbull 2018). For example, one way to calculate informativity is to use an artificial neural network (Elman 1990; Hochreiter and Schmidhuber 1997) to predict the next word given the preceding words in the utterance. The degree to which the network is surprised by the word provides an estimate of the information contributed by that word to the utterance at the point at which it occurs (*surprisal*). A word's surprisal is inversely proportional to its log probability. Averaging surprisal across all tokens of a word provides a measure of the word's informativity. Informativity is consistently found to correlate with reduction, with more informative words being less reduced, although more needs to be done to examine the relationships between the measures being examined (cf. Turnbull 2018; Zhan and Levy 2019).

[C.2] Accessibility-driven extension

Zipf (1949) found that frequent words have more senses than infrequent words, and proposed that high frequency of a form causes extension of that form to new (though similar) uses / meanings. Bybee (2003) likewise suggests that high frequency can cause extension, albeit through a different mechanism. Harmon and Kapatsinski (2017) and Koranda, Zettersten and MacDonald (2022) have shown experimentally that frequency can exert an influence on production choices: a speaker often chooses to produce a frequent form over an infrequent form even though they know the infrequent form would be a better cue to the intended meaning for a listener. In fact, it would be a better cue to the meaning for the same speaker when they themselves become a listener (Harmon and Kapatsinski 2017). These results strongly suggest that there is a chain of causation from frequency to extension. Harmon and Kapatsinski (2017) show that this chain of causation is mediated by accessibility of the form: making competing frequent and infrequent forms equally accessible eliminates the frequency effect on form choice (see also Zhan and Levy 2019 for related findings with a different paradigm). For this reason, we describe the effect as accessibility-driven extension.

Accessibility-driven extension is argued to account for the emergence of grammaticalized forms whose uses seem to defy a description in terms of necessary and sufficient semantic features. For example, Bybee and Eddington (2006) document the gradual extension of various verbs of becoming (*quedarse*, *ponerse*, *volverse*, *hacerse*) in Spanish across adjectival contexts (e.g., *quedarse solo* 'become/be left alone', *ponerse nervioso* 'become/put oneself nervous') based on semantic similarities between the adjectives. There is little that all adjectives associated with a verb share, but one can find chains of family resemblances between adjectives occurring with a verb suggesting that each verb underwent a gradual chain of extensions from adjective to adjective. Historical data support this impression (Bybee and Eddington 2006). Such complex categories of uses are difficult to account for on a bleaching account of semantic extension in which semantic features are simply lost. If bleaching was the mechanism behind such changes, we would expect all verbs of becoming to have similar contexts of use, and would not expect the contexts in which a verb is used to cluster together.

Accessibility of a form can also be influenced by other factors, like priming (Ferreira and Griffin 2003) and ease of pronunciation (Martin 2007), although frequency is often the strongest influence (e.g., Lee, Lew-Williams and Goldberg 2022). For example, the Dyirbal marker classically described as referring to women, fire and dangerous things (Lakoff 1987) has been argued to be extended to some new nouns based merely on accidental phonological similarity between the new noun and existing nouns that took the marker (Plaster and Polinsky 2011). This makes sense from an accessibility-driven extension perspective because the new noun would make the marker come to mind by activating similar nouns in memory.

Recent development of large databases of semantic changes (Zaliznjak et al. 2020) and the increasing sophistication of distributional semantic models that infer semantic similarity from text (Mikolov et al. 2017) have led to the emergence of studies that examine semantic change computationally. Some of these studies have examined the effect of word frequency in semantic change (Fugikawa et al. 2023; Hamilton, Leskovec and Jurafsky 2016; Kawasaki et al. 2022; Uban, Ciobanu and Dinu 2021; Winter and Srinivasan 2022). Hamilton, Leskovec and Jurafsky (2016) were the first to make use of distributional semantics models to examine the effects of frequency on change over time in vectors representing the positioning of words in semantic space relative to other words. They argued that infrequent words change more than frequent words. Unfortunately, these results were shown to be a mathematical artifact in Dubossarsky, Weinshall and Grossman (2017) and disappear when appropriate controls are introduced. However, Uban, Ciobanu and Dinu (2021) have used a variant of the Hamilton/Leskovec/Jurafsky method with statistical controls advocated by Dubossarsky, Weinshall and Grossman and found that frequent words change *more* than rare words over time. Kawasaki et al. (2022) took a comparative approach and compared semantic similarities (as reflected by distributional semantic models) between cognates vs. non-cognates across modern Romance languages. In contrast to Uban, Ciobanu and Dinu (2021), they found that low-frequency cognates were less similar than high-frequency cognates, and did not find a frequency effect for non-cognates, suggesting that frequent words change *less* than rare ones.

Whether these studies speak to accessibility-driven extension is questionable because not all semantic change is due to extension. Semantic changes can also be due to imperfect learning, and such changes -- like all changes due to imperfect learning -- are likelier in rare words, other things being equal. Drift due to imperfect learning probably plays a large role in semantic change because we tend to overestimate how much speakers converge on the same meaning. For example, Kapatsinski and Janda (2011) show that adult American English speakers living in the Midwest can have drastically different ideas of what states the word *Midwest* refers to, depending in part on where in the Midwest they have lived. The extent of the individual differences is such that all states except for Alaska, Maine and Hawaii were placed in the Midwest by one or more speakers: the speakers know they live in the Midwest but are not sure how far and in what directions it extends, and prioritize different dimensions. For example, a sizeable number of people living in Indiana took the Midwest to be the middle third of the United States in the North-South direction and therefore included coastal states like Oregon and Maryland. Many others included all states that were not on a coast, or included only states around the Great Lakes.

Winter and Srinivasan (2022) and Fugikawa et al. (2023) examined databases of semantic changes collected by linguists and compared the effects of concreteness and frequency on semantic change. Both studies pitted the hypothesis that frequent words are extended to infrequent meanings against the hypothesis that concrete words are extended to abstract meanings. The results were conflicting: Winter and Srinivasan observed that words with frequent words tended to be extended to less frequent meanings, with a smaller effect of concreteness, while Fugikawa et al., using a different database, found a large effect of concreteness and a small effect of frequency. Both connect their results explicitly to the hypothesis of accessibility-driven extension.

While these studies of semantic change are a promising and exciting development, a few caveats are in order. Most importantly, accessibility-driven extension does not predict a meaning frequency effect: words are often extended to meanings that are more frequent than their original meanings. For example, when a word for a body part is extended to prepositional uses, its new meaning is far more frequent than its original meaning. Rather, accessibility-driven extension predicts that, of the potential source terms in a semantic domain (e.g., body parts), the frequent ones are the ones most likely to be extended to a new domain (e.g., spatial relations).

Second, it is important to control for the relationship between the source domain and the target domain. Thus, Winter and Srinivasan (2022) show that for terms that are equally concrete and play similar functions (e.g., *skin* and *bark*), frequency plays a strong role in extension. Fugikawa et al. (2023) likewise find that the semantic relationship between source and target domain is crucial, and domain-domain mappings can vary in productivity. Accessibility-driven extension is possible only when the source form comes to mind as the speaker is trying to express a meaning in the target domain.

Accessibility-driven extension is a likely explanation for morphological paradigm leveling (Bybee and Brewer 1980; Kapatsinski 2022; Tiersma 1982). On this account of paradigm leveling, a form of a word is extended to express meanings previously expressed by other forms of the same word (Bybee and Brewer 1980). Although sometimes only a part of the word is extended, blending with the form previously expressing the meaning, this too can be due to interference from a highly accessible form during production. Evidence for the influence of accessibility was provided by Tiersma (1982) who showed that, across a wide variety of languages, the singular forms were extended to the plural *except*, crucially, for paradigms in which the plural form would be expected to be more frequent than the singular (e.g., *lice*, for which many adult native speakers of English cannot guess the singular form). In such cases, the plural form was extended to the singular. Bybee and Brewer (1980) argued that paradigm leveling tends to level minor semantic contrasts rather than more major ones (e.g., agreement vs. aspect for a verb). This too is expected on an accessibility-driven extension account because extension is caused by a form from a different paradigm cell being strongly activated by the intended meaning. When the intended meaning is highly similar to the meaning of the form, extension is more likely. Kapatsinski (2022) provides an account of these findings in a neural network framework where accessibility-driven extension is shown to be an inevitable consequence of distributed semantic representations, where the probability of extension is proportional to semantic overlap between the intended meaning and the form's established uses times the strength of association between the shared semantics and the form. The latter is at least strongly correlated with the form's token frequency.

[D] Changes that tend to affect infrequent words

Paul (1877), Wheeler (1887), Kruszewski (1890), and, more recently, Phillips (1984, 2006) proposed that some changes diffuse in the opposite direction, starting in low-frequency words and affecting high-frequency words last (if at all). A classic example is the loss of rounded front vowels in English through merger with front unrounded vowels (Phillips 1984). Changes that start in infrequent words are thought to have a different motivation, the weakness of the affected lexical representations that make them susceptible to the influence of the rest of the lexicon. For example, acoustic cues to rounding are relatively weak in front vowels: frontness raises F2, removing what is otherwise a robust cue to roundedness. Furthermore, most English front vowels are unrounded, so a listener faced with ambiguous acoustics should rationally guess that they are hearing an unrounded vowel. As a result, a listener would be relatively likely to misperceive a front rounded vowel as a front unrounded vowel. If that unrounded vowel is then taken to be part of the word's lexical representation, a sound change will have occurred in the word. Ohala (1981) was the first to propose this type of perceptual motivation for sound change. Importantly, a perceptually-motivated change is most likely to happen if the word is rare because otherwise the listener has many opportunities to correct the misperception: native listeners are quite good at their job, and so, for a native listener, a correct perception is likelier than even the likeliest of the misperceptions (Bybee 2001). An infrequent word that is misperceived is more likely to stay misperceived.

In contrast to Bybee (2001), Todd, Hay and Pierrehumbert (2019) propose that perceptually-driven sound change can be led by frequent words because an ambiguous signal is more likely to be perceived as a frequent word, and, crucially, misperceived or difficult-to-perceive tokens of a sound are less likely to be stored in memory. Together, these two assumptions mean that when a sound changes and approaches another sound, resulting in confusable, ambiguous innovative tokens, the innovative tokens of the changing sound are more likely to be stored when they come from a frequent word. This results in frequent words containing the sound changing faster than infrequent words. Todd, Hay and Pierrehumbert (2019) argue that this explains why high frequency showed a faster increase in intervocalic tapping than infrequent words in New Zealand English (as documented in Hay and Foulkes 2016). However, this proposal crucially relies on unusual, ambiguous tokens of a sound category not affecting the representation of the sound category. As far as I know, there is no evidence for this assumption, and in fact there is some evidence for the opposite, that unusual, surprising tokens of a sound category exert more influence on the sound category representation than run-of-the-mill tokens (Olejarczuk, Kapatsinski and Baayen 2018). To resort to an anecdote, 15 years after the fact, I still have a vivid memory of a token of *pen* pronounced as *pin* by a New Zealand English speaker in the ambiguous context of trying to pay with a card and being asked *Do you have a pen/pin*? When the pronunciation was (eventually) disambiguated, it was quite clear to me that I need to update my *pen* category, and a rational, error-driven listener would.

Change can also affect (some) infrequent words and leave comparable frequent ones unaffected if it is driven by *analogy* or, in Phillips' (2006) terminology, *analysis*. This is driven by the fact that the speaker has less experience with an infrequent word than with a frequent one. Therefore, when an infrequent word behaves differently from other similar words, the speaker should be less likely to be exposed to evidence that it does during language acquisition. Furthermore, even if this evidence is present in the input, there will always be less of it than there would be for a frequent word. Therefore, any rational learner should be more likely to conclude that the word is not exceptional and behaves like other similar words when the word in question is infrequent (see also Erker and Guy 2012). This is indeed what we see in learning models that allow learners to decide how a word behaves on some weighted average of experience with the word, and experience with other words (Kapatsinski 2021; Wedel 2007). The greater susceptibility of infrequent words to analogical/analytical change is the reason that old, unproductive grammatical patterns such as irregular English past tense forms survive in frequent words. These are the words for which the learner receives enough evidence to be certain of their exceptionality. Thus, *leaped* has become so common to be judged as acceptable by lexicographers, while *sleeped* and *keeped* have not (Anshen and Aronoff 1988; Hooper 1976; Lieberman et al. 2007; Song and Dalola 2019).

As mentioned earlier, many types of semantic change may also tend to affect words of low token frequency because, other things being equal, speakers are less likely to converge on the meaning of a low-frequency word. Of course, other things are often not equal because frequent words tend to have more difficult, family-resemblance category meanings due to accessibility-driven extension. Still, controlling for semantic complexity, semantic drift due to imperfect learning should affect rare words more (as observed by Kawasaki et al. 2022 through examining Romance cognates).

Rare words are also more likely to drop out of the language and be replaced by formally unrelated innovations or borrowings (Pagel, Atkinson and Meade 2007): rare words are less likely to be transmitted to the next generation. Berg (1998) and Martin (2007) showed that word loss is also affected by the same pressures that lead to sound change, i.e., phonological markedness / how difficult a word is to pronounce. As far as I know, the interaction between frequency and pronunciation difficulty has not been examined. However, it seems likely that difficult-to-pronounce frequent words are more likely to be reshaped by reductive sound change, whereas difficult-to-pronounce rare words are more likely to be replaced by unrelated borrowings.

[E] Issues and limitations

[E.1] Analogical spread of innovative reductions

Most simulations of sound change model either analogical or reductive change (e.g., Wedel 2007 for the former; Pierrehumbert 2001 for the latter). However, reductive sound change can result in a distinct way of pronouncing a sublexical unit. There is nothing in theory that would then prevent this new pronunciation variant from spreading across the lexicon by analogy. For example, flapping of /t/ and /d/ in words like *butter* and *madder* is a distinctive characteristic of American English. This is a reductive sound change that began in frequent words and is still farther along in frequent words where flaps are shorter and less constricted (Patterson and Connine 2001; Shport, Jongman and Herd 2017). It is possible to elicit full stops by presenting speakers with formal words like *emitter* but most words are probably never pronounced with a full stop in the relevant environment (following a vowel and preceding an unstressed vowel) in a typical speaker's experience. This means that novel words containing an intervocalic coronal stop are under strong analogical pressure to also be pronounced with a flap.

Kapatsinski (2021) shows that assuming an analogical pressure that is stronger in rare words and a reductive pressure that is stronger in frequent words results in a U-shaped frequency effect, whenever a reductive sound change like flapping progresses to the point that the innovative variant dominates the lexicon. At this point, novel words should fall in line with the lexicon, undergoing flapping while existing medium-frequency words can remain exceptional. To be an exception, a word needs to be associated with the conservative variant (for example, because it tends to occur in formal contexts) and to be frequent enough for its exceptionality to be learnable. It is possible for a model of change to avoid predicting a U-shaped frequency effect for an advanced reductive change, but this requires speakers to implicitly know that unknown words are likely to behave like rare words, which is not usually assumed. Exceptionally conservative behavior in rather infrequent 'learned' and formal words is often noted (e.g., Ferguson 1975), but whether such a word needs to be frequent enough to be an exception is not tested.

[E.2] Statistical issues

There is substantial uncertainty regarding whether frequency effects in corpus data should be tested using regression models with a random effect of word or if, instead, observations should be aggregated by word or even by unique values of frequency so that a random effect of word is unnecessary (Barth and Kapatsinski 2018; Gahl and Baayen 2022; Johnson 2009; Walker 2012). Barth and Kapatsinski (2018) simulated datasets with a real frequency effect and showed that a model with a random effect of word will nonetheless achieve the same fit to these data regarding of whether one includes frequency as a predictor. This is a disturbing result because we know that the frequency effect is real, yet the model does not detect it. A solution shown to work in Barth and Kapatsinski (2018) is to test the model on words it was not trained on. Gahl and Baayen (2022) advocate aggregating observations by word to avoid the same problem, although it should be noted that this is expected to reduce power. Neither solution has as yet been adopted in the literature.

Note that aggregation of observations within words is not the same as treating individual word tokens as independent but omitting a random effect of word (e.g., Erker and Guy 2012). That practice is not to be recommended as it vastly inflates significance (Barr et al. 2013; Johnson 2009; Walker 2012) and leads to poor estimates of the fixed effects (such as the effect of frequency; Barth and Kapatsinski 2018). Since most papers before 2013 took that approach, much of the early literature is on shaky statistical grounds.

Another statistical issue is that mixed-effects models assume that the levels of a random effect (such as words) come from a single normally-distributed population. As a result, when a small number of words behave exceptionally, those words are likely to be pulled in to the mean of the population of words by the model (Gelman and Hill 2007). This is especially true if the exceptional words are relatively infrequent (as in, e.g., Ferguson 1975) so the model is uncertain about their behavior and uses the rest of the lexicon to estimate it. This reduces the likelihood of detecting lexical diffusion. There are methods for detecting that random effects come from more than one population (Houghton and Kapatsinski 2023; Navarro et al. 2006) but they are not yet widely used.

Finally, a promising future direction is to make better use of hierarchical structure within the random effects by nesting collocations containing words within those words, and nesting words within word classes (Travis and Torres Cacoullos 2021). By doing so, we can disentangle the roles of specific collocations, words and their categories in lexical diffusion. Travis and Torres Cacoullos (2021) show convincing evidence of lexical diffusion of grammatical behavior by showing that specific verbs have a role above and beyond classes of verbs.

[E.3] Chaos and the actuation problem

A system is chaotic if small changes in initial conditions can lead to very different outcomes. A characteristic of a system that makes it chaotic is positive feedback loops. Language change is rife with positive feedback loops. For example, the more a form is used, the more likely it is to be reused (Martin 2007). It is therefore a good example of a chaotic system. For example, Kapatsinski (2021) shows that the simple model in Figure 1 augmented with each generation inferring a random effect of lexical identity predicts that changes will sometimes go to completion, sometimes sputter out, and usually freeze in a pattern of stable variation. The variability in outcomes is due to small stochastic noises inherent to the transmission process coupled with rich-get-richer loops. This is especially striking because the model includes no social dynamics, which are an additional source of variance in real life. Given the chaotic nature of language change, the best models of change can hope for it to describe the pressures on the system and the likely directions of change. We are unlikely to completely solve the actuation problem (Weinreich et al. 1968), predicting whether a certain change in a certain community will or will not happen. However, models of lexical diffusion can identify factors that influence likelihood of actuation. For example, the model in Kapatsinski (2021) predicts that a reductive change will be likely to sputter out if the reductive effect of frequency is relatively weak for the sound in question (either because the alternative pronunciations don't differ dramatically in difficulty or because other, social or phonological, pressures on pronunciation selection happen to be stronger), the sublexicon affected by the change happens to contain few high-frequency words, and the innovative pronunciation variant is initially rare. Sputtering out may be how changes are not actuated: innovative variants arise but then disappear because speakers infer them to have a low production probability.

[F] Conclusion

Research on frequency effects in lexical diffusion has undergone a dramatic increase in methodological sophistication in the last two decades, and continued development is to be expected in the future. The rise of large diachronic corpora, databases of changes, sophisticated models of semantics and of change itself, and experimental methods for demonstrating causal influences on change all point to an exciting time in the study of the effects of experience on both change and variable linguistic behavior more generally. There is now substantial evidence for the idea that certain types of phonetic, semantic and morphological changes are more likely to occur in frequent words. These changes are likely due to articulatory reduction -- the change is towards pronunciations characteristic of rapid and casual speech -- and accesssibility-driven lexical selection. Other changes, due to imperfect learning and analogy, are likely to occur in infrequent words and spread from there. In addition, changes accrete to words that often occur in change-favoring contexts. These results provide clear support for the central claim of usage-based linguistics -- that language structure is shaped and reshaped by language use.

Related Articles (See Also)

Article ID
WBCDL004
WBCDL008
WBCDL009
WBCDL011
WBCDL013
WBCDL017

WBCDL030
WBCDL051
WBCDL063
WBCDL068
WBCDL088
WBCDL092
WBCDL139
WBCDL140
WBCDL141

References

Alba, Matthew C. 2005. "Hiatus Resolution between Words in New Mexican Spanish: A Usage-based Account." PhD diss., The University of New Mexico.

Alba, Matthew C. 2008. "Ratio Frequency: Insights into Usage Effects on Phonological Structure from Hiatus Resolution in New Mexican Spanish." *Studies in Hispanic and Lusophone Linguistics*, 1: 247-286. DOI:10.1515/shll-2008-1020

Anshen, Frank, and Mark Aronoff. 1988. "Producing Morphologically Complex Words. *Linguistics*, 26: 641–656." DOI:10.1515/ling.1988.26.4.641

Balota, David A., Melvin J. Yap, Keith A. Hutchison, Michael J. Cortese, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. 2007. "The English Lexicon Project." *Behavior Research Methods*, 39: 445-459. DOI:10.3758/BF03193014

Baranowski, Maciej, and Danielle Turton. 2020. "TD-Deletion in British English: New Evidence for the Long-lost Morphological Effect." *Language Variation and Change*, 32: 1-23. DOI:10.1017/S0954394520000034

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal." *Journal of Memory and Language*, 68: 255-278. DOI:10.1016/j.jml.2012.11.001

Barth, Danielle. 2019. "Effects of Average and Specific Context Probability on Reduction of Function Words BE and HAVE." *Linguistics Vanguard*, 5(1). DOI:10.1515/lingvan-2018-0055

Barth, Danielle, and Vsevolod Kapatsinski. 2018. "Evaluating Logistic Mixed-effects Models of Corpus-linguistic Data in Light of Lexical Diffusion." In *Mixed-effects regression models in linguistics*, edited by Dirk Speelman, Kris Heylen and Dirk Geeraerts, 99-116. Cham: Springer. DOI:10.1007/978-3-319-69830-4_6.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Browman, Catherine, and Louis Goldstein. 1991. "Gestural Structures: Distinctiveness, Phonological Processes, and Historical Change." In *Modularity and the Motor Theory of Speech Perception: Proceedings of a Conference to Honor Alvin M. Liberman*, edited by Ignatius G. Mattingly and Michael Studdert-Kennedy, 313-338. Hillsdale, NJ: Erlbaum.

Brown, Earl K. 2020. "The Effect of Forms' Ratio of Conditioning on Word-final /s/ Voicing in Mexican Spanish." *Languages*, 5: 61. DOI:10.3390/languages5040061

Brown, Earl K., and Alba, Matthew C. 2017. "The Role of Contextual Frequency in the Articulation of Initial /f/ in Modern Spanish: The Same Effect as in the Reduction of Latin /f/? *Language Variation and Change*, 29: 57-78. DOI:10.1017/S0954394517000059

Brown, Esther L. 2004. “The Reduction of Syllable-initial /s/ in the Spanish of New Mexico and Southern Colorado: A Usage-based Approach.” PhD diss., University of New Mexico.
Brown, Esther L. 2013. “Word Classes in Phonological Variation: Conditioning Factors or Epiphenomena?” In *Selected proceedings of the 15th Hispanic Linguistics Symposium*, edited by Chad Howe, Sarah Blackwell and M. Lubbers Quesada, 179-186. Somerville, MA: Cascadilla.
Brown, Esther L., and Naomi Shin. 2022. “Acquisition of Cumulative Conditioning Effects on Words: Spanish-speaking Children’s [Subject Pronoun+Verb] Usage. *First Language* 42: 361-382. DOI:10.117/01427237211067574
Brown, Esther L., William D. Raymond, Earl K. Brown, and Richard J. File-Muriel. 2021. “Lexically Specific Accumulation in Memory of Word and Segment Speech Rates.” *Corpus Linguistics and Linguistic Theory*, 17: 625-651. DOI:10.1515/cllt-2020-0016
Brown, Esther L., and William D. Raymond. 2012. “How Discourse Context Shapes the Lexicon: Explaining the distribution of Spanish f/h words.” *Diachronica*, 29: 139–161. DOI:10.1075/dia.29.2.02bro
Bybee, Joan. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.
Bybee, Joan. 2002. “Word Frequency and Context of Use in the Lexical Diffusion of Phonetically Conditioned Sound Change.” *Language Variation and Change*, 14: 261-290. DOI:10.1017/S0954394502143018
Bybee, Joan. 2003. “Cognitive Processes in Grammaticalization.” In *The new psychology of language*, edited by Michael Tomasello, 151-174. Mahwah, NJ: Psychology Press.
Bybee, Joan. 2006. “From Usage to Grammar: The Mind's Response to Repetition.” *Language*, 82: 711-733.
Bybee, Joan. 2017. “Grammatical and Lexical Factors in Sound Change: A Usage-based Approach.” *Language Variation and Change*, *29*(3), 273-300. DOI:10.1017/S0954394517000199
Bybee, Joan L., and Mary A. Brewer. 1980. “Explanation in Morphophonemics: Changes in Provençal and Spanish Preterite Forms.” *Lingua*, 52: 201-242. DOI:10.1016/0024-3841(80)90035-2
Bybee, Joan, and David Eddington. 2006. “A Usage-based Approach to Spanish Verbs of 'Becoming'.” *Language*, 82: 323-355.
Bybee, Joan, Richard File-Muriel, and Ricardo Napoleão de Souza. 2016. “Special Reduction: A Usage-based Approach.” *Language and Cognition*, 8: 421–446. DOI:10.1017/langcog.2016.19
Bybee, Joan, and Joanne Scheibman. 1999. “The Effect of Usage on Degrees of Constituency: The Reduction of *don't* in English.” *Linguistics*, 37: 575-596. DOI:10.1515/ling.37.4.575
Cedergren, Henrietta J., and David Sankoff. 1974. “Variable Rules: Performance as a Statistical Reflection of Competence.” *Language*, 50: 333-355.
Cohen Priva, Uriel. 2017. “Not so Fast: Fast Speech Correlates with Lower Lexical and Structural Information.” *Cognition*, 160: 27-34. DOI:10.1016/j.cognition.2016.12.002
Cohen Priva, Uriel, and Emily Gleason. 2020. “The Causal Structure of Lenition: A Case for the Causal Precedence of Durational Shortening.” *Language*, 96: 413-448. DOI:10.1353/lan.2020.0025

Cohen-Goldberg, Ariel M. (2015). "Abstract and Lexically Specific Information in Sound Patterns: Evidence from /r/-Sandhi in Rhotic and Non-Rhotic Varieties of English." *Language and Speech*, 58: 522-548. DOI:10.1177/0023830914567168

De Smet, Hendrik. 2016." How Gradual Change Progresses: The Interaction between Convention and Innovation." *Language Variation and Change*, *28*: 83-102. DOI:10.1017/S0954394515000186

Dinkin, Aaron J. 2008. "The Real Effect of Word Frequency on Phonetic Variation." *University of Pennsylvania Working Papers in Linguistics*, 14: 8.

Drager, Katie, and M. Joelle Kirtley. 2016. "Awareness, Salience and Stereotypes in Exemplar-Based Models of Speech Perception and Production." In *Awareness and Control in Sociolinguistic Research*, edited by Anna M. Babel, 1-24. Cambridge: Cambridge University Press.

Dubossarsky, Haim, Daphna Weinshall, and Eitan Grossman. 2017. "Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models." In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1136-1145. Copenhagen: Association for Computational Linguistics. DOI:10.18653/V1/D17-1118.

Elman, Jeffrey L. 1990. "Finding Structure in Time." *Cognitive Science*, 14: 179-211. DOI:10.1207/S15516709cog1402_1

Erker, Daniel, And Gregory R. Guy. 2012. "The Role of Lexical Frequency in Syntactic Variability: Variable Subject Personal Pronoun Expression in Spanish." *Language*, 88: 526-557.

Ferguson, Charles A. 1975. "Short A" in Philadelphia English." In *Studies In Linguistics In Honor Of George L. Trager*, edited by M. Estellie Smith, 259-274. The Hague: Mouton.

Ferreira, Victor S., and Zenzi M. Griffin. 2003. "Phonological Influences on Lexical (Mis)selection." *Psychological Science*, 14: 86-90. DOI:10.1111/1467-9280.01424

Fidelholtz, James L. 1975. "Word Frequency and Vowel Reduction in English." *Chicago Linguistic Society*, 11(1): 200-213.

File-Muriel, Richard J. 2007. "The Role of Lexical Frequency and Phonetic Context in the Weakening of Syllable-final Lexical /s/ in the Spanish of Barranquilla, Colombia." PhD diss., Indiana University.

File-Muriel, Richard J. 2010. "Lexical Frequency as a Scalar Variable in Explaining Variation." *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 55: 1-25. DOI:10.1017/S0008413100001353

Forrest, Jon. 2017. "The Dynamic Interaction between Lexical and Contextual Frequency: A Case Study of (ING)." *Language Variation and Change*, 29: 129-156. DOI:10.1017/S0954394517000072

Fugikawa, Olivia, Oliver Hayman, Raymond Liu, Lei Yu, Thomas Brochhagen, and Yang Xu. 2023. "A Computational Analysis of Crosslinguistic Regularity in Semantic Change." *Frontiers In Communication*, 8. DOI:10.3389/Fcomm.2023.1136338.

Gahl, Susanne, and Harald Baayen. 2022. "Time and Thyme Again: Connecting Spoken Word Duration to Models of the Mental Lexicon." Ms. University Of California and University Of Tübingen.

Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Goldinger, Stephen D. 1998. “Echoes of Echoes? An Episodic Theory of Lexical Access.” *Psychological Review*, 105: 251-279. DOI:10.1037/0033-295X.105.2.251
Gregory, Michelle L., William D. Raymond, Alan Bell, Eric Fosler-Lussier and Daniel Jurafsky. 1999. “The Effects of Collocational Strength and Contextual Predictability in Lexical Production.” *Chicago Linguistic Society*, 35: 151–166.
Grondelaers, Stefan, and Stefania Marzo. 2023. "Why does the Shtyle Spread? Street Prestige Boosts the Diffusion of Urban Vernacular Features." *Language in Society*, 52: 295-320. DOI:10.1017/S0047404521001202
Hamilton, William L., Jure Leskovec, and Daniel Jurafsky. 2016. “Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change.” *Arxiv Preprint Arxiv:1605.09096*. DOI:10.48550/Arxiv.1605.09096
Harmon, Zara, and Vsevolod Kapatsinski. 2017. “Putting Old Tools to Novel Uses: The Role of Form Accessibility in Semantic Extension.” *Cognitive Psychology*, 98: 22-44. DOI:10.1016/J.Cogpsych.2017.08.002
Hashimoto, Daiki. 2019. “Cumulative Usage Effects on Lexeme-Final /s/: Probability of being Affixed.” *Lingua*, 231: 102739. DOI:10.1016/J.Lingua.2019.102739
Hashimoto, Daiki. 2021. “Probabilistic Reduction and Mental Accumulation in Japanese: Frequency, Contextual Predictability, and Average Predictability.” *Journal of Phonetics*, 87: 101061. DOI:10.1016/J.Wocn.2021.101061
Hay, Jennifer, Aaron Nolan, and Katie Drager. 2006. “From Fush to Feesh: Exemplar Priming in Speech Perception.” *The Linguistic Review*, 23: 351-379. DOI:10.1515/TLR.2006.014
Hay, Jennifer B., Janet B. Pierrehumbert, Abby J. Walker, and Patrick Lashell. 2015. “Tracking Word Frequency Effects through 130 Years of Sound Change.” *Cognition*, 139: 83–91. DOI:10.1016/J.Cognition.2015.02.012
Hay, Jennifer, and Paul Foulkes. 2016. “The Evolution of Medial /t/ over Real and Remembered Time.” *Language*, 92: 298–330.
Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. “Long Short-Term Memory.” *Neural Computation*, 9: 1735-1780. DOI:10.1162/Neco.1997.9.8.1735
Hooper, Joan B. 1976. “Word Frequency in Lexical Diffusion and the Source of Morphophonological Change.” In *Current Progress in Linguistics*, edited by William Christie, 95–105. Amsterdam: North Holland.
Houghton, Zachary N., and Vsevolod Kapatsinski. 2023. “Are your Random Effects Normal? A Simulation Study of Methods for Estimating whether Subjects or Items Come from more than one Population by Examining the Distribution of Random Effects in Mixed-Effects Logistic Regression.” *Behavior Research Methods*.
Johnson, Daniel Ezra. 2009. “Getting off the Goldvarb Standard: Introducing Rbrul for Mixed-Effects Variable Rule Analysis.” *Language and Linguistics Compass*, 3: 359-383. DOI:10.1111/J.1749-818X.2008.00108.X
Jurafsky, Daniel, Alan Bell, Michelle Gregory, and William D. Raymond. 2001. “Probabilistic Relations between Words: Evidence from Reduction in Lexical Production.” In *Frequency and the Emergence of Linguistic Structure*, edited by Joan L. Bybee and Paul J. Hopper, 229-254. Amsterdam: John Benjamins.
Kapatsinski, Vsevolod. 2009. “Adversative Conjunction Choice in Russian (*no*, *da*, *odnako*): Semantic and Syntactic Influences on Lexical Selection.” *Language Variation and Change*, 21: 157-173. DOI:10.1017/S0954394509990068

Kapatsinski, Vsevolod. 2018. *Changing Minds Changing Tools: From Learning Theory to Language Acquisition to Language Change*. Cambridge, MA: MIT Press.
Kapatsinski, Vsevolod. 2021. “Hierarchical Inference in Sound Change: Words, Sounds, and Frequency of Use.” *Frontiers In Psychology*, 12. DOI:10.3389/Fpsyg.2021.652664
Kapatsinski, Vsevolod. 2022. “Morphology in a Parallel, Distributed, Interactive Architecture of Language Production.” *Frontiers In Artificial Intelligence*, 5. DOI:10.3389/Frai.2022.803259
Kapatsinski, Vsevolod, Shelece Easterday, and Joan Bybee. 2020. “Vowel Reduction: A Usage-based Perspective.” *Rivista di Linguistica*, 32: 19-44. DOI:10.26346/1120-2726-146
Kapatsinski, V., and Richard Janda. 2011. “It’s Around Here: Residential History and the Meaning of ‘Midwest’.” *Proceedings of the Annual Meeting of the Cognitive Science Society*, 33: 2983-2988.
Kataoka, Reiko. 2011. “Phonetic and Cognitive Bases of Sound Change.” PhD diss., University of California, Berkeley.
Kawasaki, Yoshifumi, Maëlys Salingre, Marzena Karpinska, Hiroya Takamura, and Ryo Nagata. 2022. “Revisiting Statistical Laws of Semantic Shift in Romance Cognates.” In *Proceedings of the 29th International Conference on Computational Linguistics*, 141-151. Gyeongju: International Committee on Computational Linguistics.
Koranda, Mark, Martin Zettersten, And Maryellen Macdonald. 2022. “Good-enough Production: Selecting Easier Words instead of More Accurate Ones.” *Psychological Science*, 33: 1440-1451. DOI:10.1177/09567976221089603
Kruszewski, Mikołaj. 1890. “Prinzipien der Sprachentwickelung.” *Internationale Zeitschrift für Allgemeine Sprachwissenschaft*, 5: 133–144.
Kurumada, Chigusa, and Scott Grimm. 2019. “Predictability of Meaning in Grammatical Encoding: Optional Plural Marking.” *Cognition*, 191: 103953. DOI:10.1016/J.Cognition.2019.04.022
Labov, William. 1981. “Resolving the Neogrammarian Controversy.” *Language*, 57: 267-308. DOI:10.2307/413692.
Labov, William. 2020. “The Regularity of Regular Sound Change.” *Language*, 96: 42-59.
Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago, IL: University of Chicago Press.
Landauer, Thomas K., and Susan T. Dumais. 1997. “A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge.” *Psychological Review*, 104: 211-240. DOI:10.1037/0033-295X.104.2.211
Lee, Crystal, Casey Lew-Williams, and Adele Goldberg. 2022. “Accessibility Factors that Lead to Good-enough Language Production.” *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44: 2303-2309.
Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang, And Martin A. Nowak. 2007. “Quantifying the Evolutionary Dynamics of Language.” *Nature*, 449(7163): 713-716. DOI:10.1038/Nature06137
Lin, Susan, Patrice Speeter Beddor, and Andries W. Coetzee. 2014. “Gestural Reduction, Lexical Frequency, and Sound Change: A Study of Post-Vocalic /l/.” *Laboratory Phonology,* 5: 9-36. DOI:10.1515/Lp-2014-0002
Lindblom, Bjørn. 1963. “Spectrographic Study of Vowel Reduction.” *The Journal of the Acoustical Society of America*, 35: 1773-1781.

Lindblom, Bjørn, Susan Guion, Susan Hura, Seung-Jae Moon and Raquel Willerman. 1995. "Is Sound Change Adaptive?" *Rivista di Linguistica*, 7: 5-37.
Manker, Jonathan. 2019. "Contextual Predictability and Phonetic Attention." *Journal of Phonetics*, 75: 94-112. DOI:10.1016/J.Wocn.2019.05.005
Martin, Andrew T. 2007. "The Evolving Lexicon." PhD diss., University of California, Los Angeles.
McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press.
Mikolov, Tomas, Edouard Grave, Piotr Bojanowski, Christian Puhrsch, and Armand Joulin. 2017. "Advances in Pre-training Distributed Word Representations." Arxiv Preprint 1712.09405. DOI:10.48550/Arxiv.1712.09405
Mowrey, Richard, and William Pagliuca. 1995. "The Reductive Character of Articulatory Evolution." *Rivista di Linguistica*, 7: 37-124.
Navarro, Daniel J., Thomas L. Griffiths, Mark Steyvers, and Michael D. Lee. 2006. "Modeling Individual Differences Using Dirichlet Processes." *Journal Of Mathematical Psychology*, 50: 101-122. DOI:10.1016/J.Jmp.2005.11.006
Ohala, John J. 1981. "The Listener as a Source of Sound Change." In *Papers from the Parassesion on Language and Behavior*, edited by Carrie S. Masek, Roberta A. Hendrick and Mary Frances Miller, 178-203. Chicago: Chicago Linguistic Society.
Olejarczuk, Paul, Vsevolod Kapatsinski and R. Harald Baayen. 2018. "Distributional Learning is Error-Driven: The Role of Surprise in the Acquisition of Phonetic Categories." *Linguistics Vanguard*, *4*(S2). DOI:10.1515/Lingvan-2017-0020
Orozco, Rafael, and Luz Marcela Hurtado. 2020. "A Variationist Study of Subject Pronoun Expression in Medellín, Colombia." *Languages*, 6: 5. DOI:10.3390/Languages6010005
Osthoff, Hermann, and Karl Brugmann. 1878. *Morphologische Untersuchungen auf dem Gebiete der Indogermanischen Sprache*n, *I*. Leipzig: Hirzel.
Pagel, Mark, Quentin D. Atkinson, and Andrew Meade. 2007. "Frequency of Word-use Predicts Rates of Lexical Evolution throughout Indo-European History." *Nature*, *449*(7163): 717-720. DOI:10.1038/Nature06176
Patterson, David, and Cynthia M. Connine. 2001. "Variant Frequency in Flap Production." *Phonetica*, 58: 254-275. DOI:10.1159/000046178.
Paul, Hermann. 1877. "Die Vocale der Flexions- und Ableitungs-Silben in den Aeltesten Germanischen Dialecten." *Beiträge zur Geschichte der Deutschen Sprache und Literatur*, 4: 315–475.
Phillips, Betty S. 1984. "Word Frequency and the Actuation of Sound Change." *Language*, 60: 320–342.
Phillips, Betty S. 2006. *Word Frequency and Lexical Diffusion*. New York: Palgrave.
Pierrehumbert, Janet B. 2001. "Exemplar Dynamics: Word Frequency, Lenition and Contrast." In *Frequency and the Emergence of Linguistic Structure*, edited by Joan L. Bybee and Paul J. Hopper, 137–157. Amsterdam: John Benjamins.
Pierrehumbert, Janet B. 2002. "Word-specific Phonetics." *Laboratory Phonology*, 7: 101–139. DOI:10.1515/9783110197105.1.101
Plaster, Keith E., and Maria Polinsky. 2010. "Features in Categorization." In *Features: Perspectives on a Key Notion in Linguistics*, edited by Anna Kibort and Greville G. Corbett, 109-142. Oxford: Oxford University Press.

Purse, Ruaridh, Josef Fruehwald, and Meredith Tamminga. 2022. "Frequency and Morphological Complexity in Variation." *Glossa*, *7*(1). DOI:10.16995/Glossa.5839

Raymond, William D., Esther L. Brown, and Alice F. Healy. 2016. "Cumulative Context Effects and Variant Lexical Representations: Word Use and English Final T/D Deletion." *Language Variation and Change*, 28: 175-202. DOI:10.1017/S0954394516000041

Raymond, William D., and Esther L. Brown. 2012. "Are Effects of Word Frequency Effects of Context of Use? An Analysis of Initial Fricative Reduction in Spanish." In *Frequency Effects in Language Learning and Processing*, *Volume 1*, edited by Stefan Th. Gries and Dagmar Divjak, 35–52. Berlin: Mouton De Gruyter. DOI:10.1515/9783110274059.35

Schuchardt, Hugo. 1885. *Über die Lautgesetze: Gegen die Junggrammatiker.* Berlin: Oppenheim.

Shaw, Jason A., and Shigeto Kawahara. 2019. "Effects of Surprisal and Entropy on Vowel Duration in Japanese." *Language and Speech*, 62: 80-114. DOI:10.1177/0023830917737331

Shport, Irina A., Gregory Johnson, and Wendy Herd. 2017. "Flapping before a Stressed Vowel – Whatever!" *Proceedings of Meetings on Acoustics*, 31: 060004. DOI:10.1121/2.0000823

Sinclair, John. 1991. *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

Song, Jiyeon, and Amanda Dalola. 2019. "Linguistic Entrenchment and the Effect of Subjective Lexical Familiarity in Korean /n/-Insertion." *Proceedings of the Linguistic Society of America*, 4: 34-41. DOI:10.3765/Plsa.V4i1.4539

Sóskuthy, Márton, and Jennifer Hay. 2017. "Changing Word Usage Predicts Changing Word Durations in New Zealand English." *Cognition*, 166: 298-313. DOI:10.1016/J.Cognition.2017.05.032.

Sóskuthy, Márton, Paul Foulkes, Vincent Hughes, Jennifer Hay, and Bill Haddican. 2015. "Word-level Distributions and Structural Factors Codetermine GOOSE Fronting." *Proceedings of the 18th International Congress of Phonetic Sciences*.

Sosnik, Ronen, Bjoern Hauptmann, Avi Karni, And Tamar Flash. 2004. "When Practice Leads to Co-Articulation: The Evolution of Geometrically Defined Movement Primitives." *Experimental Brain Research*, 156: 422-438. DOI:10.1007/S00221-003-1799-4

Tang, Kevin, and Jason A. Shaw. 2021. "Prosody Leaks into the Memories of Words." *Cognition*, 210: 104601. DOI:10.1016/J.Cognition.2021.104601

Tiersma, Peter Meijes. 1982. "Local and General Markedness." *Language*, 58: 832-849.

Todd, Simon, Janet B. Pierrehumbert, And Jennifer Hay. 2019. "Word Frequency Effects in Sound Change as a Consequence of Perceptual Asymmetries: An Exemplar-Based Model." *Cognition*, *185*: 1-20. DOI:10.1016/J.Cognition.2019.01.004

Tomaschek, Fabian, Martijn Wieling, Denis Arnold, and R. Harald Baayen. 2013. "Word Frequency, Vowel Length and Vowel Quality in Speech Production: An EMA Study of the Importance of Experience." In *Interspeech*, 1302-1306. Lyon: International Speech Communication Association.

Traugott, Elizabeth Closs & Richard B. Dasher. 2002. *Regularity in Semantic Change.* Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs, and Ekkehard König. 1991. “The Semantics-Pragmatics of Grammaticalization Revisited.” In *Approaches to Grammaticalization*, *Volume 1*, edited by Elizabeth Closs Traugott and Bernd Heine, 189-218. Amsterdam: John Benjamins.
Travis, Catherine E., and Rena Torres Cacoullos. 2021. “Categories and Frequency: Cognition Verbs in Spanish Subject Expression.” *Languages*, *6*: 126. DOI:10.3390/Languages6030126
Turnbull, Rory. 2018. “Patterns of Probabilistic Segment Deletion/Reduction in English and Japanese.” *Linguistics Vanguard*, 4(S2). DOI:10.1515/Lingvan-2017-0033
Uban, Ana-Sabina, Alina Maria Ciobanu, and Liviu P. Dinu. 2021. “Cross-lingual Laws of Semantic Change.” In *Computational Approaches To Semantic Change*, Edited By Nina Tahmasebi, Lars Borin, Adam Jatowt, Yang Xu and Simon Hengchen, 219-260. Berlin: Language Science Press.
Walker, James A. 2012. “Form, Function, and Frequency in Phonological Variation.” *Language Variation And Change*, 24: 397-415. DOI:10.1017/S0954394512000142
Wang, William S.-Y. 1969. “Competing Sound Changes as a Cause of Residue.” *Language*, 45: 9-25.
Wedel, Andrew B. 2007. “Feedback and Regularity in the Lexicon.” *Phonology*, 24: 147-185. DOI:10.1017/S0952675707001145
Weinreich, Uriel, William Labov and Marvin Herzog. 1968. “Empirical Foundations for a Theory of Language Change.” In *Directions for Historical Linguistics*, edited by Windred Lehmann and Yakov Malkiel, 97-195. Austin: University Of Texas Press.
Wheeler, Benjamin Ide. 1887. *Analogy and the Scope of its Application in Language*. Ithaca: John Wilson and Son University Press.
Winter, Bodo, and Mahesh Srinivasan. 2022. “Why is Semantic Change Asymmetric? The Role of Concreteness and Word Frequency and Metaphor and Metonymy.” *Metaphor and Symbol*, 37: 39-54. DOI:10.1080/10926488.2021.1945419
Yun, Gwanhi. 2006. “The Effects of Lexical Frequency and Stress on Coarticulation.” *Berkeley Linguistics Society*, 32(1): 441-452.
Zalizniak, A., Smirnitskaya, A., Rousseau, M., Gruntov, I., Maisak, T., Ganenkov, D., Et Al. 2020. *Database of Semantic Shifts (2016–2020)*. Moscow: Institute of Linguistics, Russian Academy of Sciences.
Zellou, Georgia, and Meredith Tamminga. 2014. “Nasal Coarticulation Changes over Time in Philadelphia English.” *Journal of Phonetics*, 47: 18-35. DOI:10.1016/J.Wocn.2014.09.002
Zhan, Meilin, and Roger Levy. 2019. “Availability-based Production Predicts Speakers' Real-time Choices of Mandarin Classifiers.” Arxiv Preprint *Arxiv:1905.07321*. DOI: 10.48550/Arxiv.1905.07321
Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Reading, MA: Addison-Wesley.

Figure captions

Figure 1. The relationship between word frequency and reduction across 10 generations (darkest = earliest) under the assumption that each generation infers an overall probability of reduction and then produces the variants based on that probability,

frequency of the word (which has the same effect in all generations), and a constant random noise representing all other factors influencing reduction.

Figure 2. The effect of frequency by Generation. 100 grey lines are drawn for 100 replications of running the sound change through the 10 generations. Solid black line is drawn at the pre-programmed value of the frequency effect and. shows the overall (lack of) trend. The increased variability at the edges is due to difficulties of inferring the overall reduction probability when it approaches 0 and 1 because 0 and 1 correspond to minus and plus infinity on the log odds scale.

Figure 3. No reductive change happens without speakers learning to reduce at the community rate (left) or the synchronic frequency effect on reduction (right).