

Running head: FUNCTIONAL LOAD AND THE LEXICON

Functional load and the lexicon: Evidence that syntactic category and frequency relationships in  
minimal lemma pairs predict the loss of phoneme contrasts in language change

Andrew Wedel

Department of Linguistics

University of Arizona

Scott Jackson

Center for Advanced Study of Language

University of Maryland

Abby Kaplan

Department of Linguistics

University of Utah

Andrew Wedel

Department of Linguistics

University of Arizona

Tucson, Arizona 85721

email: wedel@email.arizona.edu

## ABSTRACT

All languages make systematic use of individually meaningless, contrastive categories in combination to create distinct words. Despite the central role that phonemic category contrasts play in the transmission of information, they can be lost through language change. The nearly century-old functional load hypothesis proposes that phoneme contrast loss is less likely the greater the ‘work’ that a contrast does in distinguishing words. In Wedel, Kaplan, & Jackson, under review, we showed that simple measures of functional load do significantly predict patterns of phoneme merger, and that this effect is in the hypothesized direction: the greater the number of minimal word pairs that a phoneme contrast distinguishes, the less likely is merger. Here, we extend that analysis to two additional lexical properties predicted to influence the uncertainty associated with minimal pairs. We present evidence that within our dataset, the number of minimal lemma pairs sharing syntactic category better predicts merger than those with divergent syntactic categories, and that the number of minimal lemma pairs with similar frequencies better predicts merger than those with divergent frequencies. These findings support the general variationist/usage-based/evolutionary research program, which assumes a causal chain linking properties of individual utterances to long-term change in the abstract, sublexical category system of a speech community.

# Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts in language change

## INTRODUCTION

Much theoretical work in linguistics over the last 50 years has held that the important factors accounting for sound patterns in language must be highly general and abstract (cf. the competence/performance distinction, Chomsky, 1965). Consistent with this hypothesis, many details of the pronunciation of a given phoneme token are in fact significantly predicted by its abstract phonological context, rather than by the specific word it occurs in. For example, for any word in American English, a /t/ is virtually certain to be pronounced as a flap if it occurs between two vowels, where the second vowel is unstressed. Standard phonological theories (e.g., Chomsky & Halle, 1968, Kenstowicz, 1994, Prince & Smolensky, 2004[1993]) assert that these abstract properties shared across words are the sole level that is relevant for understanding the larger phonological system of a language.

This approach has been very fruitful in describing categorical, synchronic phonological patterns within languages, but has been arguably less so in accounting for how those patterns arise and change over time e.g., (Blevins, 2004a). Further, a great deal of recent work has established that much of the variation that does exist in the pronunciation of a given phoneme is in fact predicted by a wide variety of more local, less abstract factors such as word identity (Pierrehumbert, 2002), near lexical neighbors (Wright, 2004; Munson & Solomon, 2004; Munson, 2007; Baese & Goldrick, 2009; Scarborough, 2010), and information content relative to context (Jurafsky, Bell, Gregory, & Raymond, 2001; Son & Pols, 2003; Aylett & Turk, 2004; Raymond, Dautricourt, & Hume, 2006; Kaiser, Li, & Holsinger, 2011 and many others). As a result, a

number of newer models have been developed over the last two decades that integrate various of these disparate sources of variation into accounts of the development of phonological patterns over time. These models, which we will refer to with the umbrella term ‘variationist/usage-based/evolutionary’, or VUE, differ from earlier theoretical approaches in their explicit integration of an arbitrarily large set of possible influences on pattern formation in language, including possibly ‘universal’ linguistic biases, domain-general cognitive biases as well as more idiosyncratic influences that may derive from particular structures in a language, or the culture and history of a speech community. Beyond their ability to account for a wider range of synchronic behavioral data, these models draw on a wide range of findings in fields such as evolutionary theory, cognitive science, historical linguistics and phonetics to provide new hypotheses about how variation at multiple levels and time-scales can drive the development, propagation and consolidation of phonological patterns over time (For reviews of the rationales behind this general approach, see e.g., Labov, 1994; Elman, 1995; Pierrehumbert, 2002; Blevins, 2004a; Brighton, Kirby, & Smith, 2005; Christiansen & Chater, 2008; Beckner et al., 2009; Jaeger & Tily, 2011; for examples of models, see Hare & Elman, 1995; Bybee, 2001; Blevins, 2004a; Wedel, 2006, 2007; Blevins & Wedel, 2009; Walsh, Möbius, Wade, & Schütze, 2010; Wedel, in press.)

## **The Functional Load Hypothesis**

The sound-systems of languages are constantly in flux, and over the course of time a phoneme can occasionally be lost from a language either by merging with a nearby phoneme or by eroding away in some context. As an example of phoneme merger, the historically contrastive vowels in the words *caught* and *cot* have merged for most English speakers in Canada and in the western part of the United States, with the result that these words are homophonous for these speakers. Phonemes can also be lost in specific phonological contexts, rather than lost from the entire language. An example of a context-specific phoneme loss is the case of [h] deletion before the high

front glide [j] in some east coast dialects of American English, such that the word *human* is pronounced [jumən] rather than [hjumən].

For nearly a century, a language researchers have explored the intuitively attractive idea that a phoneme contrast might be less likely to be lost from a language if it does more ‘work’ in distinguishing words in communication (Gilliéron, 1918; Trubetzkoy, 1939; Martinet, 1952; Hockett, 1967; Surendran & Niyogi, 2006; Blevins & Wedel, 2009; Silverman, 2010; Kaplan, 2011; Wedel et al., under review). The notion of ‘work’ in this sense has been operationalized and investigated in a number of ways, but until recently, clear evidence in favor of this *functional load* hypothesis has been elusive. In Wedel et al. (under review), we provided the first statistical evidence in support of the functional load hypothesis by showing that within a database drawn from 8 languages, phoneme contrasts that distinguish more minimal pairs were significantly less likely to have merged over the course of time. Furthermore, for those phoneme contrasts that distinguish few or no minimal pairs, we found that phoneme frequency was positively correlated with merger probability. As described below, these two findings are consistent with VUE models of language change.

In this paper, we present new evidence consistent with the functional load hypothesis that sound change is biased toward selective maintenance of those phonemes that specifically contribute to reducing uncertainty between competing lexical items in discourse (Blevins & Wedel, 2009; Wedel, in press). This finding represents a challenge to the classic model of phonology, in which particularities of existing words and their usage patterns can have no systematic influence on phoneme pronunciation. As a result, classic models cannot easily link sound change to contrast among actual lexical items; indeed, generative phonological formalisms are explicitly designed to exclude actual words from consideration (Padgett, 2003; see discussion section below). In contrast, VUE models provide the mechanistic tools for such a link as well as corresponding testable hypotheses (Wedel, in press).

## Model Predictions

The VUE family of models has a long history reaching back to the 19th-century work of (Courtenay, 1895). Modern VUE models draw on findings showing that mental categories maintain some record of experienced variation, rather than consisting solely of abstract generalizations (reviewed in the context of language in Johnson, 1997; Bybee, 2002; Pierrehumbert, 2002; Baayen, 2007; Pisoni & Levi, 2007; Ernestus, 2011). Further, experiencing category token variants has been shown to influence subsequent production (Goldinger, 2000; see also Pardo, 2006 and Nielsen, 2007 and perception (Dennis Norris, 2003; Eisner & McQueen, 2005; Kraljic & Samuel, 2005a, 2005b) of tokens of the same and similar categories (reviewed in Pierrehumbert, 2002, 2003; Wedel, in press). This body of findings suggests that a speaker's lexical category system can be understood as a steadily-updating multi-dimensional network, in which experienced phonetic detail can be represented at multiple levels of analysis and where generalizations can emerge from and coexist with that detail (see, e.g., Elman, 1995; Bybee & McClelland, 2005; Beckner et al., 2009; Walsh et al., 2010; Wedel, in press). This conception of the lexicon as a densely interconnected network of representations of different granularities provides a pathway by which variants can spread from word to word e.g., (Wang, 1969; Bybee, 2002; Phillips, 2006) and from sound to sound (e.g., Kraljic & Samuel, 2005a; Mielke, 2008) over lifetimes and generations within a speech community (Wedel, in press).

Within this general type of model, any consistent biases on variation in production or perception can feed into this network to promote systematic change in a language system over time. Researchers using VUE models tend not to place *a priori* limits on what the sources of linguistic variation may be, and argue that many crosslinguistically common patterns arise because of common biases in articulation, perception and cognition shared among humans (Blevins, 2004a, see also Christiansen & Chater, 2008). At the same time, language-specific biases that arise, for example, from idiosyncratic structural features of a language, sociolinguistic processes, or language contact are also predicted to play roles in shaping particular pathways of

language change. As a result, VUE models predict that a fuller account of the origin of phonological patterns in language will require reference to at least some of the particular properties and contexts of individual languages (see Labov, 1994, 2001; Blevins, 2004a for arguments in favor of this position, and Blevins, 2004b; Chitoran & Hualde, 2007; Blevins, 2009 for some particular examples.).

VUE models contribute testable hypotheses about the relationship between the role of phoneme contrast in information transmission and the probability of phoneme contrast loss. More specifically, (Wedel, 2004, 2006, in press) describes and simulates a multi-level exemplar model (Walsh et al., 2010) for a general linking mechanism between biased variation in word production/perception events and long-term change in phonetic distributions within sound categories. This Network-Feedback model argues that given evidence for the storage of experienced detail on the one hand and production/perception feedback on the other, any mechanism in production or perception favoring phonetically more contrastive tokens of minimal pair members will promote maintenance of a phonetic distinction between the phonemes defining that minimal pair across the lexicon. This model is based in part on the many studies that have reported that in production, phonetic cues to word identity are enhanced when a word has more near neighbors, that is, when there are more words in the lexicon that are pronounced similarly (Aylett & Turk, 2004; Wright, 2004; Munson & Solomon, 2004; Munson, 2007; Baese & Goldrick, 2009; Scarborough, 2010). Interestingly, Baese and Goldrick (2009) report that a phoneme contrast distinguishing a minimal pair is relatively hyperarticulated in productions of one member of the pair even when the other is not present in the context. One interpretation of this finding is that a speaker's aggregate experience of lexical uncertainty and associated cue-enhancement in usage influences the longer term representation of phonetic detail in lexical categories (for additional arguments in favor of this conclusion, see also (Cohen Priva, 2008)). Additionally, a wide range of experimental work has shown that cues to word identity tend to be enhanced/reduced when the word is less/more predictable given the local discourse or sentential

context, e.g., (Jurafsky et al., 2001; Son & Pols, 2003; Aylett & Turk, 2004; Raymond et al., 2006; Cohen Priva, 2008; Kaiser et al., 2011) . These findings together suggest that enhancement of phonetic cues to a word’s identity may be correlated with the actual and expected degree of uncertainty in categorization for that word in usage.

A consequence of these findings for a the Network-Feedback model is that merger probability for a given phoneme pair should depend on both how often and to what degree a phoneme contrast reduces uncertainty associated with a minimal pair. In Wedel et al. (under review), we identified two factors related to the frequency of lexical uncertainty: the number of lexical minimal pairs distinguished by the phoneme pair in question, and in the case that there were few or no minimal pairs, the frequency of the relevant phonemes <sup>1</sup>. The minimal pair result suggests that loss of a phoneme contrast is inhibited in relation to the *number* of minimal pairs distinguished by that contrast, or framed in terms of the model, the number of minimal pairs for which uncertainty is strongly dependent on that phoneme contrast. Here we test two additional factors that are available to us in our dataset that are predicted to be correlated with the *degree* of uncertainty associated with any minimal pair.

- Minimal pairs that share the same syntactic category (e.g., noun-noun, verb-verb) should be associated with greater uncertainty, and therefore inhibit merger more than minimal pairs with different categories (e.g., noun-verb). The uncertainty associated with minimal pairs that share syntactic category should be greater because local morpho-syntactic context will contribute less to their disambiguation, leaving a correspondingly greater role for phonetic cues.

- Minimal pair members with similar frequencies should inhibit merger more strongly than than those with divergent frequencies, because *ceteris paribus*, uncertainty is greater between two system elements of similar probabilities (Shannon, 1948).

Within our datasets we found that both of these predictions are tentatively confirmed. In the following sections, we describe the construction of the datasets in more detail, followed by analysis and discussion.



# CORPUS STUDY

## Database

The rate of phoneme merger over the course of language change tends to be low, with the result that often only a small number of historically recent phoneme mergers are attested in related variants of any given language. Consequently, in order to obtain enough data for statistical analysis, we pooled data from multiple languages. The languages represented in the dataset are English (Received Pronunciation and Standard American), German, Dutch, French, Spanish, Turkish, Korean, and Hong Kong Cantonese.

The dataset consists of 18 systems of phoneme contrasts from these 8 different languages; these groups are summarized in Table 1. Each system of phoneme contrasts consists of phonemes within a single structural class such as ‘vowels’ or ‘consonants’, and the the set of phoneme contrasts within each group is limited to those differing in only one phonological feature such as *voice* or *place of articulation*. Each system contains at least one phoneme contrast that has merged in some dialect of the language, as well as all other phonologically similar phoneme contrasts in that structural class. Each system, then, can be considered as a comparison within a language of a set of phoneme contrasts that have merged and a set of structurally similar phoneme contrasts that have not. In total, the dataset contains 56 phoneme contrasts that have merged, and 524 that have not. Within mergers, 35 are consonant-consonant mergers as opposed to vowel-vowel mergers, and 30 are conditioned by phonological context, as opposed to context-free. A context-free merger eliminates a phoneme contrast from a language altogether, while a context-sensitive one eliminates a contrast only in certain phonological environments. The North American English *cot*  $\sim$  *caught* merger is an example of a context-free merger, while the *pin*  $\sim$  *pen* merger, which occurs in southern dialects of American English, merges [ɪ]  $\sim$  [ɛ] only before nasals. As a consequence, the words *pin* and *pen* are homophonous in these dialects, while *pit* and *pet* are not. In addition to phoneme – phoneme category mergers, our dataset contains

three phoneme deletions which we model as merger with zero.

Word and frequency information for each language is obtained from a phonemically-transcribed corpus. No grammatical or function words were included. These corpora are different on a number of dimensions (e.g., size and source genre), as are the languages they represent (e.g., complexity of phoneme inventory, syllable structure, complexity of morphology). As described below, we used hierarchical modeling to model the differences between languages as random effects. The fact that these models represent “partial pooling” of results across the languages makes our results more generalizable than simple logistic regression (Gelman & Hill, 2007) and therefore should be fairly robust against the heterogeneous nature of the languages and corpora

For all the languages in this study except for Korean and Hong Kong Cantonese, corresponding surface-form and lemmatized versions of each individual corpus were available. For the analysis comparing the role of word-category in the minimal pair count factor (described below), we used a dataset which excluded Korean and Hong Kong Cantonese so that every phoneme contrast in the dataset could be associated with both lemma- and surface form-based minimal pair counts. A major difference between lemma and surface-form minimal pair counts lies in the fact that for a lemma-based count, each root participates in at most one minimal pair with another root, while in the surface-form based count, each word that a root participates in can contribute an independent minimal pair. As an example using the minimal root pair *pit* and *pat* in English, verbal suffixes create the four minimal word pairs, *pit*  $\sim$  *pat*, *pits*  $\sim$  *pats*, *pitting*  $\sim$  *patting*, and *pitted*  $\sim$  *patted*. In the lemma-based count, there would be one minimal pair for this root pair, while in the surface-form-based count there would be four. Consequently, the resulting minimal pair count based on surface-forms is a mix in which some roots contribute many minimal pairs to the count and others only one.

The Korean corpus was available in a lemmatized form, and although the Hong Kong Cantonese corpus is technically a surface-form corpus, we treated it as a lemmatized corpus

because of the isolating morphology of the language. For the subsequent analysis investigating the role of minimal pair member frequency relationships, we used only lemma-based minimal pair counts allowing us to include the data from Korean and Hong Kong Cantonese.

## Analysis

In order to address the question of which formulations of functional load are most appropriate, we will use the framework of logistic regression, by fitting models to predict cases of phoneme merger. More specifically, because the data is structured hierarchically, we will use hierarchical generalized linear models (also known as logistic mixed-effects models), fitted using the “lme4” package for the R statistical software (Bates et al citation, R Core Working Group citation). This approach should provide more robust generalization across languages, avoiding both the overfitting that would result from regular logistic regression and the loss of power if models were fit separately for each language in our database. In the models that follow, *phoneme system* is used as the grouping variable (or random effects variable). This means that each system (e.g., American English consonant-consonant mergers, American English vowel-vowel mergers before nasals, tone mergers in Hong Kong Cantonese) is modeled with a random intercept, allowing the models to account for the fact that mergers happen to be more common in some systems. Random slopes – which allow the effects of our variables of interest to also vary randomly by system – were also fit, but these models were generally not better-fitting than the simpler models, and did not provide any different conclusions. We make note of any exceptions to this in the text.

Within this regression-based framework, we are faced with the challenge of comparing the predictive utility of highly-correlated variables. The distribution of phonemes across the lexicon is largely, though not entirely random (Tamariz, 2008), and as a consequence, most measures relating phonemes to words will correlate with one another. Therefore, teasing apart which more

---

<sup>0</sup>The CMU pronouncing dictionary lacks frequency information. Frequencies for this dataset were taken from CELEX.

nuanced formulations of functional load are more effective predictors is a challenging task. We will pursue two types of analysis, and if they both provide corroborating evidence, we will take that as initial support for choosing one predictor over another, even if the predictors are highly correlated. The two analytic techniques are (1) model comparison using model fit statistics and likelihood ratio tests of nested models, and (2) comparisons of residualized versions of the predictors.

In the first step, models with the competing predictors are fit, and the model fit statistics of AIC and BIC are compared. This allows us to make an initial, tentative claim on which predictor is superior. Then both competing models are compared with a “superset” model that contains both competing predictors, using a likelihood ratio test. If these tests show that the superset model is significantly better than the model with the worse predictor, but not significantly better than the model with the better predictor, this is taken as further evidence that the ‘better’ predictor really is better.

In order to complement this analysis, we also residualize both of the competing predictors in turn, using a simple linear model. Then we fit two models, one with the ‘real’ predictor A and the residualized predictor B, and one with the real predictor B and the residualized predictor A. The first model allows us to see whether the variance in B not predicted by A adds any significant prediction above and beyond A, and the second model allows us to investigate the converse. If the model comparison step above selected predictor A, and we found evidence that the residualized predictor A added prediction above predictor B, but the residualized B did not add above predictor A, then we would have further evidence that predictor A is superior to predictor B.

These two complementary techniques are used in the following section to test the relationship of phoneme merger probability to lexical factors that are predicted to influence uncertainty between minimal pair members.

## RESULTS

### Role of Roots and Their Syntactic Category in Merger Probability

Wedel et al., under review, report that the more minimal pairs distinguished by a phoneme contrast, the less likely that contrast is to merge. We first address the issue of whether within-category (e.g., noun  $\sim$  noun) minimal pairs are more important for driving the effect than between-category (e.g., noun  $\sim$  verb) minimal pairs. We will simultaneously address a second question, whether minimal pair counts should be computed using lemma-based counts, or surface-form based counts. Both lemmas and surface-forms are *a priori* plausible influences on merger probability. On the one hand, surface-forms more closely represent the tokens that are actually encountered in speech, and thus could better reflect the potential ambiguities that speakers may encounter. Silverman (2010) argues on this basis that surface-forms, and not lemmas, should be used in evaluating the functional load hypothesis. On the other hand, we are unaware of any research that directly addresses how relationships among affixal morphemes influence the cue enhancement effects reviewed in the introduction above, but to the extent that they does not, counting minimal pairs by surface-form rather than lemma introduces potentially extraneous information reflecting morphological paradigm structure in the lexicon, rather than just content-morpheme relationships. Our results show that compared to the alternatives, the count of within-category, lemma-based minimal pairs is the most effective as a predictor of merger.

We start by taking the subset of data for which we have both lemma-based and surface-form counts of minimal pairs (all languages but Korean and Hong Kong Cantonese; see the 'Database' section above). This comprises a total of 35 mergers across 482 phoneme contrasts in 7 corpora. We then calculate within- and between-category minimal pair counts, for both lemmas and surface-forms, and we re-arrange the data so that 'lemma-based' is treated as nominal 'predictor', distinguishing the two types of counts for each of the phoneme pairs. We can then fit the overall model from Wedel et al., under review, which includes segment token

frequency and its interaction with the presence/absence of (any) minimal pairs, in addition to the measures of functional load under investigation here. We fit this full model, because ultimately we are interested in the best predictors within the context of this overall model. To start, we fit a model that includes both within- and between-category minimal pair counts, and we model the interaction of these variables with our dummy ‘lemma-based’ variable.<sup>2</sup> This complex model provides the following insight: for surface-based counts, between-category minimal pairs are a significant predictor ( $\beta = -2.22$ ,  $z = -3.31$ ,  $p = 0.001$ ), but within-category are not ( $\beta = 0.21$ ,  $z = 0.59$ ,  $p = 0.555$ ), but both of these interact significantly with the lemma-based dummy predictor ( $\beta = 3.51$ ,  $z = 2.01$ ,  $p = 0.044$  and  $\beta = -9.05$ ,  $z = -2.7$ ,  $p = 0.007$ , respectively), which indicates that this pattern is reversed for lemma-based counts.

In order to see this cross-over more clearly, we fit separate models for lemma-based counts and for surface-based counts. The lemma-based model shows that within-category minimal pairs is a significant predictor ( $\beta = -3.38$ ,  $z = -2.4$ ,  $p = 0.016$ ), while between-category minimal pairs is not ( $\beta = 0.92$ ,  $z = 0.88$ ,  $p = 0.38$ ). Conversely, the model using surface-based counts shows that within-category pairs is not significant as a predictor ( $\beta = 0.29$ ,  $z = 0.6$ ,  $p = 0.551$ ), but between-category pairs is ( $\beta = -2.67$ ,  $z = -3.06$ ,  $p = 0.002$ ). In order to choose from these two possibilities, we provide both empirical and theoretical arguments in favor of the within-category lemma-based counts.

The empirical argument is provided by model comparison and residualization techniques. At the simplest level, a model including within-category lemma-based counts as the measure of functional load has better fit statistics than a model using between-category surface-form counts (AIC = 235.22 vs. 239.4, BIC = 260.29 vs. 264.47). When these models are compared to a superset model with both factors using likelihood ratio tests, the superset model is not significantly better than the model employing within-category lemma-based counts ( $\chi^2 = 0.15$ ,  $df = 1$ ,  $p = 0.699$ ), but it is significantly better than the model using between-category surface-form counts ( $\chi^2 = 4.33$ ,  $df = 1$ ,  $p = 0.037$ ). Finally, residualization of the two measures

shows that a residualized within-category lemma-based counts still predicts merger significantly above and beyond the information provided by between-category surface-form counts ( $\beta = -1.08$ ,  $z = -2.03$ ,  $p = 0.043$ ), but that a residualized between-category surface-form counts does not add prediction above the information provided by within-category lemma-based counts ( $\beta = -0.16$ ,  $z = -0.33$ ,  $p = 0.74$ ). In summary, both the model comparison and residualization techniques indicate that within-category lemma-based minimal pair counts is a more effective predictor of phoneme merger than between-category surface-form counts. It is furthermore important to note that the between-category surface-form measure is actually more correlated with within-category *lemma*-based counts than it is with the within-category *surface*-based counts is (0.86 vs. 0.79). We therefore assume that the apparent effectiveness of between-category surface-form counts, in particular why it appears to be better than within-category surface-form counts, is that it is simply more correlated with the overall most effective predictor, within-category lemma-based counts.

The theoretical argument reduces to an argument of parsimony. From its inception, the functional load insight has related phoneme-merger probability to the role of phoneme contrast in supporting transmission of information in usage. The wealth of current data we have reviewed above supports this view, and is consistent with a model that frames functional load in terms of reduction of uncertainty at the level of individual utterances. From this standpoint, syntactic category is clearly predicted to play a role in the relationship of minimal pairs to merger probability, since local syntactic and/or morphological context should often strongly reduce the uncertainty between minimal pair members of differing syntactic category. We might imagine some alternative model for a relationship between minimal pair counts and phoneme merger probability that is not related to information transmission, but it is not clear why such a model would predict that between-category minimal pairs should have a greater influence on phoneme merger probability than within-category minimal pairs. Following these empirical and theoretical arguments, in the remaining analyses we will concentrate only on lemma-based counts of

within-category minimal pairs.

## Role of Frequency Ratio

A further prediction of a VUE model of sound change is that the *relative* frequencies of the members of a minimal pair should influence the average degree of uncertainty between them<sup>3</sup>. Specifically, all else being equal, uncertainty is expected to be greater between minimal pairs members whose frequencies are similar (Shannon 1948). We can conceptualize this influence of relative frequencies as a weight assigned to each minimal pair corresponding to the modulation of the expected uncertainty by the difference in frequencies. While there are many ways to operationalize this prediction, one straightforward way is to use the ratio of frequencies as a weighting factor in the minimal pair count. To do this, we calculated a simple ratio of the frequency of the lower frequency minimal pair member divided by the frequency of the higher frequency member, and summed these footnote/Using this approach, minimal pair members with similar frequencies will have a ratio near 1, while those with very divergent frequencies will have a ratio closer to zero. We have also tested the log-transformation of this frequency ratio, as well as the ratio of the unigram surprisals (i.e., the ratio of the base-2 logarithm of the frequency of the minimal pair member in the corpus), but these produce essentially the same pattern of results described in the text.. However, in our data, we observed an extremely high correlation of ( $r = 0.985$ ,  $N = 580$ ) between this weighted measure and the simple count of (within-category, lemma-based) minimal pairs. This high correlation is perhaps unsurprising if the distribution of minimal lemma pairs result from processes that mimic two random draws from the set of lemmas with their associated frequencies.

Despite this very high correlation, we investigated whether model comparison could tease apart the predictive power of the ratio-weighted minimal pair counts described above from simple minimal pair counts.. The model-fit statistics were extremely close (AIC = 325.07 vs. 323.38, BIC = 351.25 vs. 349.55), and neither model with each measure as the lone functional load variable



were significantly different from the superset model including both (simple minimal pair count:  $\chi^2 = 1.7$ ,  $df = 1$ ,  $p = 0.193$ ; frequency ratio:  $\chi^2 = 0$ ,  $df = 1$ ,  $p = 0.961$ ). As a consequence, this formulation of a predicted frequencies ratio effect does not provide evidence either way.

However, it is also plausible that the relationship between frequency ratio and functional load of a minimal pair is non-linear. Therefore, we investigated a different formulation, capturing the notion that minimal pairs with relatively similar frequencies should contribute more to functional load than those with dissimilar frequencies. For this measure, we calculated a median minimal-pair frequency ratio independently for each of the 18 phoneme systems in the dataset and then counted the number of minimal pairs for each phoneme pair whose ratio was above or below the corresponding median ratio. This provides a simple way to split the minimal pair count into two bins, one containing minimal pairs with more balanced frequencies (i.e., having a ratio closer to 1) and the other containing minimal pairs with more divergent frequencies (i.e., having a ratio closer to zero). As expected, these two measures, which we will call *balanced* and *unbalanced* minimal pair counts, are also highly correlated. Nevertheless, our results provide some evidence that the balanced minimal pair count is a significantly better predictor of the probability of merger within our dataset.

As in our analyses so far, the first step is model comparison. The model with the balanced minimal pair variable has somewhat better fit statistics than the model with the unbalanced minimal pair variable (AIC = 322.02 vs. 329.27, BIC = 348.19 vs. 355.45). This difference is confirmed by likelihood ratio tests, which show that the superset model with both predictors significantly outperforms the model with the unbalanced minimal pair variable ( $\chi^2 = 8.46$ ,  $df = 1$ ,  $p = 0.004$ ), but does not significantly outperform the model with the balanced minimal pair variable ( $\chi^2 = 1.21$ ,  $df = 1$ ,  $p = 0.272$ ). Finally, the comparison of residualized variables leads to the same conclusion. When the balanced minimal pair variable is residualized on the unbalanced minimal pair variable, it still acts as a significant predictor alongside the unbalanced minimal pair variable ( $\beta = -0.89$ ,  $z = -2.68$ ,  $p = 0.007$ ), but the same is not true of the unbalanced minimal

pair variable; when residualized on the balanced minimal pair variable, it no longer adds prediction above the balanced minimal pair variable ( $\beta = 0.31$ ,  $z = 1.11$ ,  $p = 0.267$ ). In summary, both the model-comparison and residualization techniques indicate that there is enough distinct information in these highly correlated variables to suggest that the count of “balanced” minimal pairs (i.e., the minimal pairs where the probabilities of the members are relatively close) matters more to the prediction of merger than the “unbalanced” minimal pairs (i.e., minimal pairs where the members have very different probabilities).

While these techniques converge on the same conclusion, we should be cautious for two reasons. First, because the two measures are so highly correlated, their significant difference in predictive power could be due to a small number of datapoints. Second, this effect is only obtained when we make a fairly arbitrary decision in how to formulate the measure. This suggests that if the observed effect is not an artifact of some aspect of our data, there is some kind of non-linearity in the relationship between frequencies ratio and functional load. This nature of this possible non-linearity remains to be explored, and raises additional questions which may not be able to be adequately addressed with the data presented here.

In order to partially address the first concern, we can display the relationship between the fitted values of the two competing models, to get a sense of how or why the balanced minimal pair variable is outperforming the unbalanced minimal pair variable. The fitted values are the estimated probability of merger, as predicted by each of the models. Figure 1 displays the differences between the fitted values on the y-axis, plotted as a function of the values predicted by the balanced minimal pair model. Points above the diagonal line indicate that the balanced minimal pair model predicted greater probability of merger for that phoneme contrast. Mergers are plotted as large crosses, and non-mergers are plotted as small circles. The models are performing very similarly overall, which is expected due to the high correlations between the predictors. However, note that at the upper right of the graph where most points represent mergers, the balanced model generally makes stronger predictions in favor of merger than the

unbalanced model. Conversely, in the lower left where most points represent non-mergers, the balanced model generally makes stronger predictions in favor of non-merger than the unbalanced model. The generality of this pattern suggests that the better performance of the balanced model is not due to a handful of extreme points.

This pattern can be seen more easily if these fitted values are plotted in a different way. Figure 2 plots the differences between the fitted values (which is equivalent to the vertical distance between points in figure 1 and the slope = 1 line) as a function of the fitted values for the balanced minimal pair model. That is, the values on the x-axis in figure 2 represent the probability of merger for each phoneme contrast in the data set, as predicted by the balanced minimal pair model. Values on the y-axis are the difference in fitted values between the balanced and unbalanced minimal pair models. Greater vertical distance above the dotted line indicating  $y = 0$  indicates that the balanced model predicts greater probability of merger than the unbalanced model, and conversely, greater distance below the dotted line indicate that the balanced model predicts lower probability of merger than the unbalanced model. To the extent that the superiority of the balanced model is broadly-based in the data, we expect a consistent divergence between the two models particularly where the balanced model makes its most extreme predictions. This appears to be the case: where the balanced model more strongly predicts non-merger (corresponding to low values along the x-axis), most points represent non-merged phoneme contrasts, and their position below the dotted-line indicates that the balanced model predicts non-merger more strongly. Similarly, where the balanced model more strongly predicts merger, most points do represent mergers and most tend to be above the dotted line, indicating that for these phoneme contrasts the balanced model predicts merger more strongly than the unbalanced model. Again, the fact that we see this general trend throughout the data suggests that the better performance of the balanced model in predicting merger probability in our dataset is not due to a few isolated cases.

We can use these fitted values to find example cases for which the balanced minimal pair model is doing a better job. The shaded areas in Figure 2 show quadrants where the balanced minimal pair model is clearly outperforming the alternative, as described above. These regions cover 5 cases of merger and 8 cases of non-merger, which are given as examples in table 3. Perhaps the most important thing to note from these examples is that the model is making these stronger predictions for a range of languages and phoneme classes in the dataset, suggesting that model performance in this dataset is not based on an artifact of inclusion of a particular language.

## Summary of Final Model

We have presented arguments for a more specific formulation of the functional load variable, namely a lemma-based count minimal pairs, where the minimal pair members are of the same syntactic category, and have unigram probabilities that are above the median for the phoneme system. Model comparison and residualization techniques provided statistical evidence that this formulation is a more effective predictor of merger than other reasonable alternatives. In this section, we provide a statistical and graphical summary of the full model. Following the initial results described by Wedel et al., under review, the model consists of a variable representing functional load, and a variable representing segmental (token) frequency which interacts with a dichotomized functional load variable.<sup>4</sup> In table 4 we give the coefficients for the parameters in the final model.<sup>5</sup> The pattern of effects follows the same pattern as that found in Wedel et al., under review. First, where there are any (within-category) minimal pairs, there is a significant effect of functional load, here specified as the number of within-category minimal pairs that have relatively balanced probability ratios (i.e., relatively close to 1), compared to the median frequency ratio value for the phoneme system. This effect is in the expected direction, such that the greater the functional load, the less likely merger is to occur. Second, for phoneme contrasts with minimal pairs, the effect of segment frequency is non-significant. Third, for phoneme

contrasts with *no* minimal pairs, the effect is significantly more positive, such that greater segment frequency leads to greater probability of merger.

While this model is simple and provides significant prediction, a great many other factors contribute to actual patterns of sound change (see e.g. Labov, 1994, 2001; Blevins, 2004a). Nonetheless it is interesting to ask how well this simple model performs in separating mergers from non-mergers. This can be displayed graphically by inspecting the distribution of predicted values (i.e., fitted values) of merger probability separately for actual mergers and non-mergers. These distributions are shown for our data in the density plots in figure 3. This figure shows that despite the extreme simplicity of this model, separation is fairly good. Predicted probabilities of merger somewhere above 0.30 appear to characterize mergers almost exclusively, and the majority of non-mergers are assigned merger probabilities of 0.10 or lower. The general point here is that for being so simple, the model is doing rather well given the complexities of historical sound change.

One might also wonder to what extent these results are expected to generalize to languages not investigated here. It is certainly the case that the language-specific estimates of overall rate of merger contribute a great deal to the predicted probabilities shown in figure 3. However, because we have not found any evidence that the fixed-effect coefficients (i.e., the effects in table 4) vary significantly across phoneme systems, we can have reasonable expectations that the general pattern of results would generalize, and that probabilities of merger assigned by the model would be successful in assigning relative merger probabilities to a similar degree as seen in the present data. However, given that the rate of merger is typically quite low for any given language, we cannot expect any given language to show statistically significant evidence of a functional load effect on its own. In fact, we expect that evidence supporting a functional load effect has proven elusive for many years precisely because sufficient linguistic and computational resources have been lacking to study this until recently.

## DISCUSSION

### Functional Load and Traditional Generative Phonology

We have shown that the likelihood of diachronic merger between two sounds depends on the amount of ‘work’ that the pair does in distinguishing words in the lexicon. Furthermore, the evidence indicates that this effect cannot be explained solely as a phoneme-level phenomenon (driven, for example, only by phoneme frequency): the functional load of a phoneme pair also depends on the actual words that it distinguishes in the lexicon.

This result stands in striking contrast to the approach of traditional generative phonology, which makes a sharp distinction between *possible* words (knowledge of which is part of a speaker’s phonological grammar) and *actual* words (which are an accident of the lexicon). Theories that explicitly aim to handle contrast, such as Dispersion Theory (Flemming, 2002; Campos-Astorkiza, 2007; Ní Chiosáin & Padgett, 2009; Padgett, 2009) and various implementations of underspecification (e.g., Itô & Mester, 1986), Hall, 2011, deal with the level of the segment or, at most, idealized possible words. Formal devices for maintaining contrast between actual words are typically limited to morphological paradigms, in the form of constraints such as ANTI-IDENT (Crosswhite, 1999), REALIZEMORPH (Kurusu, 2001), PARADIGMCONTRAST (Kenstowicz, 2002; Itô & Mester, 2004), or PRESERVECONTRAST applied to members of a paradigm (Ouwayda, 2010)

This focus on possible as opposed to actual words is a deliberate choice, meant to capture the fact that speakers *do* distinguish between nonwords that are phonotactically legal in their language and those that are not. In addition, giving the phonological grammar unrestricted access to the lexicon has the potential to predict unattested patterns:

The idea of neutralization avoidance, if understood in the wrong way, can make strange predictions. For example, consider the fact that Standard English has the words *beat* [bit], *boot* [but], and *peat* [pit], but no *poot* [put]. If [i] and [u] are

unmarked because they make a perceptually good contrast, and [i] is even better in the absence of such a contrast, then do we expect [pit] to become [pit] (since there is no [put] for [pit] to remain distinct from)? Similarly, if there were a process backing [i] to [u], would we expect that it might affect [pit] but not [bit], since only the latter would entail a neutralization (with [but])?

These questions arise when we take the domain of explanation to be the set of actual lexical items in a language. But this is in fact not the practice in generative phonology. Instead, theories model the set of *possible* words of a language.... (Padgett, 2003, 78-79)

This generalization appears to be correct: we are aware of only one pattern that has been argued to involve synchronic homophony avoidance among paradigmatically unrelated forms (Ichimura, 2006); however, as Mondon (2009, 157-9) notes, the evidence presented by Ichimura leaves open the possibility that the pattern is not productive.

The approach of traditional generative phonology, then, is a reasonable and principled one. However, in eliminating their ability to describe *synchronic* homophony avoidance, these theories also give up any possibility of describing the *diachronic* effects of functional load as well. We emphasize again that it would not be possible to capture these diachronic effects by making a simple modification to the machinery of a typical generative theory – for example, by ‘decorating’ a language’s phoneme inventory with information about the frequency of each segment. Our results suggest that the functional load effect depends on specific relationships among actual lexical items, which is precisely what traditional generative phonology rejects.

One response to this state of affairs is to conclude that functional load effects are located outside the regular phonological machinery. However, both the control of pronunciation details in usage and the generation of systematic phoneme-inventory patterns are central phonological phenomena. Restricting the domain of ‘phonology’ to exclude lexical properties that have been shown to influence these phenomena would be an unwarranted rejection of potential explanatory

power. Thus, it is worth exploring VUE models that unite the synchronic variation that is the traditional purview of phonology with the diachronic effects seen here.

## **Functional Load and Variationist/Usage-Based/Evolutionary Theories**

In stark contrast to classic models, a central feature of VUE models is the assumption of a causal chain linking properties of individual utterances to long-term change in the abstract, linguistic category system of a speech community (Ohala, 1989; Labov, 1994; Kirby, 1999; Bybee, 2001; Pierrehumbert, 2001, 2003; Blevins, 2004a; Wedel, 2007; Christiansen & Chater, 2008; Beckner et al., 2009; Blevins & Wedel, 2009). This approach predicts that any consistent bias in utterances that can be perceived and reproduced by language users can in principle influence the trajectory of language change. The functional load hypothesis represents a potentially fruitful place to look for evidence for or against such a long-range connection, because it explicitly relates change in a phonological system to actual usage events of words in a language.

Specifically, the functional load model predicts that the utility of a phoneme contrast in reducing uncertainty between words in usage influences the likelihood of loss of that phoneme contrast through language change. Part of the historic difficulty in finding evidence for such a link lies in the fact that functional load is only one of many influences on sound change, with the result that statistical techniques need to be applied to a relatively large dataset to find evidence for it. An additional problem is that many of the possible factors related to a functional load effect are very tightly correlated with one another, as expected if the distribution of phonemes across words is largely random. As a consequence, a correspondingly large amount of data is required to exploit the few deviations that reveal tighter association of a given factor with merger probability than another. Wedel et al. (under review) showed that a simple count of minimal pairs distinguished by a phoneme contrast significantly predicted merger of that contrast in some community speaking that language. Phoneme frequency was also found to be predictive of merger, but in the opposite direction: higher phoneme frequency is correlated with merger, but



also correlated with higher minimal pair count. Interestingly in this regard, the influence of phoneme frequency was only significantly predictive of merger in the absence of minimal pairs. Nonetheless, because of the strong correlation between minimal pair counts and other non-lexical factors, we need to be cautious about accepting that features of the lexicon influence sound change in this way. As a further test of the involvement of the lexicon in the functional load effect, in the work reported here we investigated two additional, specifically lexical factors that are predicted to influence lexical uncertainty in usage within minimal pairs: syntactic category, and relative minimal pair frequency. Our finding that including both factors in the analysis improves the model is consistent with the hypothesis that actual words matter.

### **Syntactic-Category as a Factor in Merger Prediction**

In Wedel et al. (under review), we showed that for a given phoneme contrast in a language in our dataset, the number of words distinguished by that contrast inversely correlates with merger probability in that language. Here, we divided the minimal pair-count factor into two sub-categories on the basis of whether the minimal pair members were of the same syntactic category or different syntactic categories. The rationale for this division is that uncertainty for within-category minimal pair members should be on average greater because local morpho-syntactic context is less likely to contribute to their disambiguation. In addition, each of the corpora that underlie the dataset are available in lemmatized and surface-form versions, so we created corresponding lemma- and surface-form-based counts for within- and between-category minimal pairs. Consistent with prediction, we found that the lemma-based count of within-category minimal pairs predicted merger significantly better than any of the other three minimal pair counts. Residualization of these other three counts against the lemma-based, within-category count suggests that they derive their predictive power from their correlation with the lemma-based, within-category count.

Counting minimal pairs is a very coarse technique for estimating the causal factor in the

functional load effect that is predicted by model, that is, the overall reduction in uncertainty in language use contributed by a particular phoneme contrast. In this regard, it is interesting that the within-category lemma-based count is more predictive than the corresponding surface-form based count. This suggests that for this coarse measure, at least, the additional information about morphological paradigm structure inherent in the surface-form count is not relevant.

## **Frequency Ratio as a Factor in Merger Prediction**

The relative frequency of within-category, lemma-based minimal pair members was also found to be a significant predictor of merger, consistent with prediction. To the extent that predictability of words in context is related to their unigram frequency of occurrence, the uncertainty between minimal pair members should be on average greater when their unigram frequencies are similar (Shannon, 1948). If cue-enhancement is related to the degree of uncertainty, minimal pairs for which the members are more similar in frequency should contribute more to the inhibition of merger. We found that when the within-category minimal pair count for each phoneme pair was split by the median low/high frequency ratio for minimal pairs within a phoneme system, the count of minimal pairs with closer frequencies predicted merger significantly better than the count of those with more divergent frequencies. Consistent with the notion that this effect derives from the specific minimal pair member frequencies rather than some correlated, more general feature of the lexicon, corresponding factors for the between-category minimal pair count did not show a difference in predictive power. Several exemplar-based models of phonetic category merger also show more probable and more rapid merger when phonetically adjacent categories are produced with divergent frequencies (Pierrehumbert, 2001; Wedel, in press)

Finally, we note that we did not find any measure based on absolute word frequency that was predictive of merger probability within this dataset. There are a number of plausible, non-exclusive explanations for this. First, the dataset is based on modern corpora which will fail in any number of ways to accurately reflect the state of the language at the inception and

progress of the merger. We expect that the frequency relationships within the corpus are likely to be less reflective of the relevant state of the language than the more categorical measures of word existence and word-category. Second, unigram frequency itself may simply be a relatively poor predictor of uncertainty relevant to phonetic cue enhancement processes (see e.g., (Piantadosi, Tily, & Gibson, 2011)).

## **Lexical-Level Entropy as a Factor in Merger Prediction**

Hockett (1967), followed by Surendran and Niyogi (2003, 2006), note that the change in lexical-level entropy upon loss of a phoneme contrast represents that contrast’s contribution to the overall information transmission capacity of a system (Shannon, 1948) and argue that it is therefore likely to be the best measure of functional load. In Wedel et al. (under review), and here for datasets based both on lemma and surface-form frequencies, we find that the more local measures of minimal pair count and relative phoneme frequency do a significantly better job of predicting merger. As noted above, it is possible that the frequency information in the corpora on which the dataset is based may not correspond sufficiently well to the language context during the merger process to reveal an underlying relationship with lexical-level entropy. However, we think that the lexical-level entropy contribution of a phoneme pair is unlikely to have a direct influence on merger probability for the following reasons. If cue enhancement in phonemes were directly related to lexical-level entropy contribution, we should see equivalent degrees of enhancement in every instance of each phoneme without regard to its local uncertainty in context. Instead, as mentioned above, research has established that contextual uncertainty has a significant influence on pronunciation variation. It is true that evidence also suggests that some more global degree of expected uncertainty influences pronunciation separately from the properties of the immediate context (Baese & Goldrick, 2009; Cohen Priva, 2008). However, it is more parsimonious to explain this broader behavior solely in terms of the observed (e.g., Wang, 1969; Bybee, 2002; Phillips, 2006; Mielke, 2008), and experimentally supported (e.g., Kraljic &

Samuel, 2005a) 'bottom-up' generalization of initially context-specific pronunciation variants as proposed in the Network Feedback model (Wedel, 2004, in press), rather than in addition through some top-down influence of lexical-level entropy. It is worthwhile noting, however, that bottom-up generalization of phoneme variants which reduce uncertainty between specific lexical items will indirectly act to support the lexical-level information-transmission capacity of the system as a whole.

## CONCLUSIONS

In addition to providing statistical evidence for a link between the probability of phoneme contrast loss and the role of phoneme contrast in information transmission in language use, this study provides some guidance for future work on the topic. Where previous literature has operationalized the notion of *functional load* in a wide variety of ways, our data suggests that some measures capture the relevant phenomena better than others, in particular that lemma-based counts are more predictive of merger than surface-form-based counts, and that local measures such as relative frequency within minimal pairs are more predictive than more global measures such as system entropy (cf. Surendran & Niyogi, 2003). These findings suggest two clear avenues for further research: (i) enlarging the database to improve our ability to separate highly correlated variables, in particular by including more languages with different properties and from different language families; and (ii) exploring more direct methods to measure the uncertainty between minimal pair members, for example by comparing the n-gram conditional probabilities of minimal pair members in speech corpora.

However, despite the simplicity of this model, it may already contribute to our understanding of certain phonological patterns. As a matter of expositional convenience, up to this point we have primarily framed the model in terms of predicting the probability of phoneme

contrast loss, but it can just as well be thought of as making predictions about the probability of phoneme contrast preservation. As an example, consider the unusually crowded high/mid/front vowel space of American English, containing the contrastive vowels  $/i \sim \text{ɪ} \sim \text{e} \sim \text{e}/$ . Within a genetically- and areally-balanced sample of 628 languages (Mielke, 2008), only 6% of the languages' vowel inventories include this set of four vowels. As a comparison, 69% of the languages' vowel inventories include the more dispersed high/mid set  $/i \sim \text{e} \sim \text{u} \sim \text{o}/$ . Given the relative typological rarity of  $/i \sim \text{ɪ} \sim \text{e} \sim \text{e}/$  as a set, and the observation that they are more confusable with each other than with other vowels in the American English system (Hillenbrand, Getty, Clark, & Wheeler, 1995), we might expect that some of these contrasts would fully merge in some natively English speaking speech communities. Instead, these vowel contrasts often chain-shift with respect to each other in ways that preserve their distinctiveness (see Labov, 1994; MacLagan & Hay, 19 for examples). In 4 we repeat the figure comparing model predictions for mergers and non-mergers, indicating the model predictions for the context-free mergers of  $/i \sim \text{ɪ}/$ ,  $/\text{ɪ} \sim \text{e}/$  and  $/\text{e} \sim \text{e}/$  vowel contrasts, with the actually merged  $/\text{a} \sim \text{o}/$  contrast included for comparison. We see that the model makes its strongest possible predictions that the phonetically similar high-mid-front vowel pairs in American English will *not* merge across-the-board, while being essentially agnostic about the  $/\text{a} \sim \text{o}/$  contrast. This supports hypotheses that the failure to find context-free mergers of these high/mid/front vowel contrasts in dialects of English is related to their high functional load (e.g., Labov, 1994; MacLagan & Hay, 19). This is consistent with the Network-Feedback model, which describes a mechanism for a high functional load for a phoneme contrast to promote chain-shifting or contrast trading over merger; for a detailed description of this model and illustrative simulations, see Wedel, in press. Finally, in the introduction we noted that in addition to biases that are likely to be common among all humans, VUE models predict that idiosyncratic structural properties of languages should also play a significant role in shaping pathways of language change. The evidence we report here is consistent with this prediction, suggesting that the particular distribution of phonemes across the *actual* lexicon of a language

will contribute significantly to an account of the evolution of its phoneme inventory.

## REFERENCES

- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and speech*, 47, 31-56.
- Baayen, R. H. (2007). Storage and computation in the mental lexicon. In G. Jarema & G. Libben (Eds.), *The mental lexicon: Core perspectives* (p. 81-104). Philadelphia: Elsevier.
- Baayen, R. H., Piepenbrock, R., & Rijn, H. van. (1995). *The CELEX lexical database*. Philadelphia: Linguistic Data Consortium. (Release 2, CD-ROM.)
- Baese, M., & Goldrick, M. (2009). Mechanisms of interaction in speech production. *Language and cognitive processes*, 24, 527-554.
- Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system: Position paper. *Language learning*, 59(s1), 1-26.
- Blevins, J. (2004a). *Evolutionary phonology*. Cambridge: Cambridge University Press.
- Blevins, J. (2004b). The mystery of austronesian final consonant loss. *Oceanic Linguistics*, 43, 179-184.
- Blevins, J. (2009). Structure-preserving sound change: A look at unstressed vowel syncope in austronesian. In A. Adelaar & A. Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for bob blust* (p. 33-49). Canberra: Pacific Linguistics.
- Blevins, J., & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica*, 26, 143-183.
- Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In M. Tallerman (Ed.), *Language origins: Perspectives on evolution*. Oxford: Oxford University Press.
- Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically

- conditioned sound change. *Language Variation and Change*, 14, 261-290.
- Bybee, J., & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22, 381-410.
- Campos-Astorkiza, J. R. (2007). *Minimal contrast and the phonology-phonetics interaction*. Unpublished doctoral dissertation, University of Southern California.
- Chitoran, I., & Hualde, J. (2007). From hiatus to diphthong: The evolution of vowel sequences in romance. *Phonology*, 24, 37-75.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.
- Christiansen, M. N., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31, 489-509.
- Cohen Priva, U. (2008). Using information content to predict phone deletion. In N. Abner & J. Bishop (Eds.), *Proceedings of the 27th west coast conference on formal linguistics* (p. 90-98). Somerville, MA: Cascadilla Proceedings Project.
- Courtenay, J. Baudoin de. (1895). *Versuch einer Theorie phontischer Alternationen: Ein Capitel aus der Psychophonetik*. Strassburg, Germany: Trüber.
- Crosswhite, K. (1999). Intra-paradigmatic homophony avoidance in two dialects of Slavic. In M. K. Gordon (Ed.), *Papers in phonology 2* (Vol. 1, p. 48-67). Los Angeles: UCLA Department of Linguistics.
- Dennis Norris, A. C., James McQueen. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204-238.
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67, 224-238.
- Elman, J. L. (1995). Language as a dynamical system. In R. Port & T. van Gelder (Eds.), *Mind as motion: Dynamical perspectives on behavior and cognition*. Cambridge, MA: MIT Press.



- Ernestus, M. (2011). Categoricality and gradience in phonology theory. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology*. Oxford: Blackwell.
- Fagyal, Z., Kibbee, D., & Jenkins, F. (2006). *French: A linguistic introduction*. Cambridge: Cambridge University Press.
- Flemming, E. S. (2002). *Auditory representations in phonology*. New York, NY: Routledge.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Gilliéron, J. (1918). *Généalogie des mots qui désignent l'abeille d'après l'atlas linguistique de la france*. Paris: É. Champion.
- Goldinger, S. D. (2000). The role of perceptual episodes in lexical processing. In *Spoken word access processes* (p. 155-158). ISCA.
- Hall, D. C. (2011). Phonological contrast and its phonetic enhancement: Dispersedness without dispersion. *Phonology*, 28, 1-54.
- Hare, M., & Elman, J. L. (1995). Learning and morphological change. *Cognition*, 56, 61-98.
- Harris, J. (1969). *Spanish phonology*. Cambridge, MA: MIT Press.
- Hillenbrand, J., Getty, L., Clark, M., & Wheeler, K. (1995). Acoustic characteristics of american english vowels. *J. Acoust. Soc. Am*, 97(5), 3099-3111.
- Hockett, C. (1967). The quantification of functional load. *Word*, 23, 320-339.
- Ichimura, L. K. (2006). *Anti-homophony blocking and its productivity in transparadigmatic relations*. Unpublished doctoral dissertation, Boston University, Boston, MA.
- Itô, J., & Mester, A. (1986). The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry*, 17, 49-73.
- Itô, J., & Mester, A. (2004). Morphological contrast and merger: *Ranuki* in Japanese. *Journal of Japanese Linguistics*, 20, 1-18.
- Jaeger, T. F., & Tily, H. (2011). Language processing complexity and communicative efficiency.

- WIRE: Cognitive Science*, 2, 323-335.
- Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (p. 145-165). San Diego, CA: Academic Press.
- Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (p. 229-254). Amsterdam: John Benjamins.
- Kaiser, E., Li, D., & Holsinger, E. (2011). Exploring the lexical and acoustic consequences of referential predictability. In I. Hendrickx, A. Branco, L. Devi, & R. Mitkov (Eds.), *Anaphora processing and applications* (Vol. 7099, p. 171-183). Cambridge, MA: Springer.
- Kang Kwong, L. (2004). *Hong Kong Cantonese corpus*. Available from <http://www0.hku.hk/hkcancor/>.
- Kaplan, A. (2011). How much homophony is normal? *Journal of linguistics*, 47, 631-671.
- Kenstowicz, M. (1994). *Phonology in generative grammar*. Cambridge, MA: Blackwell.
- Kenstowicz, M. (2002). Paradigmatic uniformity and contrast. In A. Csirmaz, Z. Li, A. Nevins, O. Vaysman, & M. Wagner (Eds.), *Phonological answers (and their corresponding questions)* (Vol. 42, p. 141-163). Cambridge, MA: MIT Working Papers in Linguistics.
- Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.
- Kissine, M., Velde, H. van de, & Hout, R. van. (2003). An acoustic study of standard Dutch /v/, /f/, /s/ and /z/. In L. Cornips & P. Fikkert (Eds.), *Linguistics in the Netherlands* (p. 93-104). Amsterdam: John Benjamins.
- Kraljic, R., & Samuel, A. G. (2005a). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review*, 13, 262-268.
- Kraljic, R., & Samuel, A. G. (2005b). Perceptual learning for speech: Is there a return to normal?

- Cognitive Psychology*, 51, 141-178.
- Kurisu, K. (2001). *The phonology of morpheme realization*. Unpublished doctoral dissertation, University of California, Santa Cruz.
- Labov, W. (1994). *Principles of linguistic change* (Vol. I: Internal Factors). Oxford: Blackwell.
- Labov, W. (2001). *Principles of linguistic change* (Vol. II: Social Factors). Oxford: Blackwell.
- Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.
- Lee, Y. (2006). *Sub-syllabic constituency in Korean and English*. Unpublished doctoral dissertation, Northwestern University.
- Lewis, G. (1967). *Turkish grammar*. Oxford University Press.
- MacLagan, M., & Hay, J. (19). Getting fed up with our feet : Contrast maintenance and the new zealand english ‘short’ front vowel getting fed up with our feet : Contrast maintenance and the new zealand english ‘short’ front vowel shift. *Language Variation and Change*, 1(1-25).
- Martinet, A. (1952). Function, structure, and sound change. *Word*, 8, 1-32.
- Mendonça, A., Graff, D., & DiPersio, D. (2009). *Spanish gigaword second edition*. Linguistic Data Consortium, Philadelphia.
- Mielke, J. (2008). *The emergence of distinctive features*. Oxford: Oxford University Press.
- Mok, P. M.-K., & Wong, P. W.-Y. (2010). Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. In *Speech prosody 2010, fifth international conference*. Chicago: International Speech Communication Association.
- Mondon, J.-F. (2009). *The nature of homophony and its effects on diachrony and synchrony*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia, PA.
- Munson, B. (2007). Lexical access, lexical representation, and vowel production. In J. S. Cole & J. I. Hualde (Eds.), *Papers in laboratory phonology* (Vol. IX, p. 201-228). New York: Mouton de Gruyter.
- Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel

- articulation. *Journal of speech, language, and hearing research*, 47, 1048-1058.
- New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du Français contemporain sur internet: LEXIQUE. *L'année psychologique*, 101, 447-462.
- Ní Chiosáin, M., & Padgett, J. (2009). Contrast, comparison sets, and the perceptual space. In S. Parker (Ed.), *Phonological argumentation: Essays on evidence and motivation* (chap. 4). London: Equinox.
- Nielsen, K. (2007). The interaction between spontaneous imitation and linguistic knowledge. *UCLA Working Papers in Phonetics*, 105, 125-137.
- Ohala, J. (1989). Sound change is drawn from a pool of synchronic variation. In L. E. Breibik & E. H. Jahr (Eds.), *Language change: Contributions to the study of its causes*. Berlin: Mouton de Gruyter.
- Ouwayda, S. (2010, January). *Contrast preservation in dialects of North Levantine Arabic*. (Poster presented at the 84th Annual Meeting of the Linguistic Society of America in Baltimore, MD)
- Padgett, J. (2003). Contrast and post-velar fronting in Russian. *Natural Language and Linguistic Theory*, 21, 39-87.
- Padgett, J. (2009). Systemic contrast and Catalan rhotics. *The Linguistic Review*, 26, 431-463.
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382-2393.
- Penny, R. J. (2002). *A history of the Spanish language* (2nd ed.). Cambridge: Cambridge University Press.
- Phillips, B. (2006). *Word frequency and lexical diffusion*. New York: Palgrave MacMillan.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9), 3526-3529.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure*

- (p. 137-157). Amsterdam: John Benjamins.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory phonology* (Vol. VII, p. 101-139). Berlin: Mouton de Gruyter.
- Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, 46, 115-154.
- Pisoni, D., & Levi, S. (2007). Some observations on representations and representational specificity in speech perception and spoken word recognition. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (p. 3-18). Oxford: Oxford University Press.
- Prince, A., & Smolensky, P. (2004[1993]). *Optimality Theory: Constraint interaction in generative grammar*. Oxford: Blackwell.
- Raymond, W., Dautricourt, R., & Hume, E. (2006). Word-medial /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18, 55-97.
- Sak, H., Güngör, T., & Saraçlar, M. (2008). Turkish language resources: Morphological parser, morphological disambiguator and web corpus. *Proceedings of GoTAL*, 5221, 417-427.
- Scarborough, R. (2010). Lexical and conceptual predictability: Confluent effects on the production of vowels. In C. Fougerson, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.), *Papers in laboratory phonology* (Vol. 10, p. 557-586). Berlin: de Gruyter.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System technical journal*, 27, 379-423, 623-656.
- Silverman, D. (2010). Neutralization and anti-homophony in Korean. *Journal of linguistics*, 46, 453-482.
- Sohn, H.-H. (1999). *The Korean language*. Cambridge: Cambridge University Press.
- Son, R. J. J. H. van, & Pols, L. C. W. (2003). How efficient is speech? In E. H. Berkman (Ed.), *Proceedings of the institute of phonetic sciences*. Amsterdam.
- Surendran, D., & Niyogi, P. (2003). *Measuring the functional load of phonological contrasts* (Tech.

- Rep. No. TR-2003-12). Chicago: Department of Computer Science, University of Chicago.
- Surendran, D., & Niyogi, P. (2006). Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In O. N. Thomsen (Ed.), *Competing models of linguistic change: Evolution and beyond* (p. 43-58). Amsterdam: John Benjamins.
- Tamariz, M. (2008). Exploring systematicity between phonological and context-cooccurrence representations of the mental lexicon. *The Mental Lexicon*, 3, 259-278.
- Trubetzkoy, N. (1939). *Grundzüge der phonologie*. Prague, Czech Republic: Travaux du Cercle Linguistique de Prague.
- Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive science*, 34, 537-582.
- Wang, W. S. Y. (1969). Competing changes as a cause of residue. *Language*, 45, 9-25.
- Wedel, A. (2004). *Self-organization and categorical behavior in phonology*. Unpublished doctoral dissertation, University of California, Santa Cruz.
- Wedel, A. (2006). Exemplar models, evolution and language change. *The linguistic review*, 23, 247-274.
- Wedel, A. (2007). Feedback and regularity in the lexicon. *Phonology*, 24, 147-185.
- Wedel, A. (in press). Lexical contrast maintenance and the development of sublexical contrast systems. *Language and Cognition*, VOLUME, PAGES.
- Wedel, A., Kaplan, A., & Jackson, S. (under review). High functional load inhibits phonological contrast loss: A corpus study. *Cognition*.
- Weide, R. L. (1995). *The CMU pronouncing dictionary*. Carnegie Mellon University. Available from <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>.
- Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.
- Wiese, R. (2000). *The phonology of German*. Oxford: Oxford University Press.
- Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology* (Vol. VI, p. 26-50). Cambridge:

Cambridge University Press.

Zee, E. (1985). Sound change in syllable final nasal consonants. *Journal of Chinese linguistics*, 13, 291-330.

Zee, E. (1999). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese linguistics*, 27, 120-167.

## **AUTHOR NOTE**

The authors would like to thank T. Florian Jaeger and Jared Linck for useful discussion. All remaining mistakes remain the sole responsibility of the authors.



## FOOTNOTES

<sup>1</sup>We hypothesize that in the absence of any functional load effect, the probability of phoneme contrast merger is a function of the probability that the distributions of two phonemes drift toward one another sufficiently that they can be reanalyzed as one category by a subsequent generation. Within some VUE models, more frequent use has been predicted to correlate with more rapid change because each use provides an opportunity for the production of new variation. As a result, ‘apparent time’ may run faster for more frequently used sound categories. (See e.g., Bybee, 2001 and many others).

<sup>2</sup>As mentioned above, we fit a variety of random-effects models, including the full random slopes model. All of these models produced the same pattern of effects reported here, but the simplest model – the model with only a random intercept effect of phoneme system – was found to be the best fitting. It is the results from this model that are reported here.

<sup>3</sup>Because all frequencies reported in the text are normalized by dividing by the corresponding total corpus frequency, they correspond to unigram probabilities.

<sup>4</sup>See Wedel et al., under review for a full discussion of this model and the justification of the dichotomized interaction.

<sup>5</sup>This model includes a random intercept term for phoneme system, as described below, but the random intercept values are not given here. More complex random effects did not improve the fit of the model and also did not change the pattern of results for the fixed effects.

Language	Phoneme Types / Context	Actual Mergers	Notes	References
English ( <i>RP</i> )	V ~ V	aɪ ~ ɔɪ ʊə ~ ɔɪ ɪə ~ ɛə ɛɪ ~ ɛə	PRICE ~ CHOICE CURE ~ THOUGHT NEAR ~ SQUARE NURSE ~ SQUARE	Wells (1982)
	C ~ C	θ ~ t θ ~ f θ ~ s		
English ( <i>American</i> )	V ~ V	a ~ ɔ ɔɪ ~ ɛɪ	LOT ~ THOUGHT	Labov et al. (2006)
	V ~ V / __ ɪ	a ~ ɔ	START ~ NORTH	
	V ~ V / __ n	ɪ ~ ɛ	PIN ~ PEN	
	V ~ V / __ l	ɪ ~ i	HILL ~ HEEL	
		ʊ ~ u	PULL ~ POOL	
		ʊ ~ oʊ	BULL ~ BOWL	
		ʌ ~ a	HULL ~ HALL	
		ʊ ~ ʌ	BULL ~ HULL	
	C ~ C	ʌ ~ w		
	C ~ Ø / __ j	j ~ Ø		Labov et al. (2006)
German	V ~ V	e ~ ɛɪ	GEBE ~ GÄBE	Wiese (2000)
Dutch	C ~ C	s ~ z f ~ v x ~ ɣ		Kissine et al. (2003)
French	V ~ V	ẽ ~ œ e ~ ɛ œ ~ ø ~ ɔ	VIN ~ UN ÉPÉE ~ ÉPAIS	Fagyal et al. (2006)
Spanish	C ~ C	ʎ ~ j s ~ θ		Penny (2002, 106) Harris (1969)
Turkish	C ~ Ø / V __ V	ɣ ~ Ø		Lewis (1967)
Korean	C ~ C / [-son] __	p ~ p' t ~ t' s ~ s' tʃ ~ tʃ' k ~ k'	post-obs. tensing	Sohn (1999, 173)
	C ~ C / __ ]σ	p ~ p <sup>h</sup> t ~ t <sup>h</sup> ~ s ~ s' ~ tʃ ~ tʃ <sup>h</sup> ~ h k ~ k <sup>h</sup> ~ k'	coda neut.	Sohn (1999, 165)
Cantonese ( <i>Hong Kong</i> )	C ~ C / # __	n ~ l		Zee (1999)
	C ~ C / __ #	n ~ ŋ		Zee (1985)
	C ~ Ø / __ ]σ	ŋ ~ Ø		Zee (1999)
	T ~ T	2 ~ 5		Mok & Wong (2010)

Table 1

*Phoneme pairs and known actual mergers included in the database.*

Language	Corpus	surface-forms Available?
English ( <i>RP</i> )	CELEX (Baayen, Piepenbrock, & Rijn, 1995)	yes
English ( <i>American</i> )	CMUDict <sup>1</sup> (Weide, 1995)	yes
German	CELEX (Baayen et al., 1995)	yes
Dutch	CELEX (Baayen et al., 1995)	yes
French	Lexique (New, Pallier, Ferrand, & Matos, 2001)	yes
Spanish	Spanish Gigaword Corpus (Mendonça, Graff, & DiPersio, 2009)	yes
Turkish	(Sak, Güngör, & Saraçlar, 2008)	yes
Korean	Korean National Database (Lee, 2006)	no
Cantonese ( <i>Hong Kong</i> )	Hong Kong Cantonese Corpus (Kang Kwong, 2004)	no

Table 2

*Corpora used to construct the database.*

Corpus	System	Pair	Merged	Balanced Count	Unbalanced Count
Fr	F_VV	ɛe	Merged	9	10
Fr	F_VV	øɔ	Merged	2	6
Ko	K_3_CodaNeut	ʈʈ	Merged	1	4
Ko	K_3_CodaNeut	s t	Merged	1	2
RP	RP_CC	v ■	Merged	0	1
Am	AE_VV_1	e i	Unmerged	19	10
Du	D_CC	f x	Unmerged	46	25
Du	D_CC	j ʏ	Unmerged	21	14
Fr	F_VV	e ɔ	Unmerged	35	28
Ko	K_5_OnsNeut	ʈʈʰ	Unmerged	14	5
RP	RP_CC	b v	Unmerged	26	15
RP	RP_CC	f v	Unmerged	25	17
RP	RP_CC	h ʈʈ	Unmerged	21	15

Table 3

*Selected examples predicted especially well by balanced minimal pair model*

parameter	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-3.20	0.45	-7.17	0.0000
functional load (special min. pair count)	-3.11	0.83	-3.74	0.0002
absence of minimal pairs	-0.47	0.45	-1.04	0.2965
segment frequency	0.31	0.23	1.34	0.1793
segment freq. x absence of min. pairs	1.21	0.52	2.31	0.0206

Table 4

*Fixed-effect parameters for final model*

## FIGURE CAPTIONS

*Figure 1.* Scatterplot of fitted values for different frequency ratio variables

*Figure 2.* Differences of fitted values between two models

*Figure 3.* Distribution of predicted probabilities for final model

*Figure 4.* The position of several American English vowel contrasts within the distribution of predicted probabilities for final model









