Brief article

# High functional load inhibits phonological contrast loss: A corpus study

Andrew Wedel [a],*, Abby Kaplan [b], Scott Jackson [c]

[a] University of Arizona, Dept. of Linguistics, Douglass Bldg., Rm. 200E, P.O. Box 210028, Tucson, AZ 85721, United States
[b] University of Utah, Dept. of Linguistics, Langs. & Comm. Bldg., Rm. 2300, 255 S Central Campus Dr., Salt Lake City, UT 84112, United States
[c] University of Maryland, Center for Advanced Study of Language, 7005 52nd Ave., College Park, MD 20742, United States

## ARTICLE INFO

## ABSTRACT

For nearly a century, linguists have suggested that diachronic merger is less likely between phonemes with a high functional load – that is, phonemes that distinguish many words in the language in question. However, limitations in data and computational power have made assessing this hypothesis difficult. Here we present the first larger-scale study of the functional load hypothesis, using data from sound changes in a diverse set of languages. Our results support the functional load hypothesis: phoneme pairs undergoing merger distinguish significantly fewer minimal pairs in the lexicon than unmerged phoneme pairs. Furthermore, we show that higher phoneme probability is positively correlated with merger, but that this effect is stronger for phonemes that distinguish no minimal pairs. Finally, within our dataset we find that minimal pair count and phoneme probability better predict merger than change in system entropy at the lexical or phoneme level.

© 2013 Elsevier B.V. All rights reserved.

## 1. Introduction

Spoken languages make use of a system of individually meaningless, contrastive sound categories, often termed *phonemes*, in combination to create distinctive words (Hockett, 1960; Studdert-Kennedy & Goldstein, 2003). Despite the central role phonemes play in carrying contrast between words, phonemes can be lost from a language when, for example, two phonemes merge with one another (Labov, 1994, ch. 11). For example, in many regions of North America the historically contrastive vowels in the words *cot* and *caught* have merged, with the result that these words are now homophonous. Nearly a century ago, Gilliéron (1918) first proposed that the probability of phoneme loss should be inversely related to the amount of 'work' that the phoneme does in distinguishing words in communication. Termed the *functional load hypothesis* by Jakobson (1931), Mathesius (1931), and Trubetzkoy (1939), and developed further by Martinet (1952) and

Hockett (1967), the idea that change in a system of phonemes is related to their role in information transmission has held great intuitive appeal for language-change researchers over the last century. However, clear evidence supporting this hypothesis has not been found. Previous work has focused on individual case studies due to limited access to data and computational resources, and results of these studies have been inconclusive (Blevins & Wedel, 2009; King, 1967; Kaplan, 2011; Silverman, 2010; Surendran & Niyogi, 2006). This is perhaps not surprising: even if functional load does influence the probability of phoneme loss, many other systemic and phonetic (Blevins, 2004; Labov, 1994) as well as social (Labov, 2001) factors also influence sound change. As a consequence, we would expect to find many individual 'exceptions' to the functional load hypothesis even if functional load can contribute significantly to the course of sound change. Testing the functional load hypothesis requires a larger sample of data in which effects can be assessed statistically.

In this paper, we present the first such analysis of a dataset comprising a large number of phoneme mergers from a diverse set of languages. We show for the first time

---

that simple measures of functional load within a system of phonemes *do* significantly predict patterns of phoneme merger, and that this effect is in the hypothesized direction: the greater the contribution a pair of phonemes makes to word differentiation, the less likely those phonemes are to merge over the course of language change. Further, we show that in the case that a phoneme pair does not distinguish many words, phoneme probability is a significant predictor of merger.

## 2. Corpus study

### 2.1. 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), Korean, French, German, Dutch, Slovak, Spanish, and Hong Kong Cantonese. Each language is represented by a phonemically-transcribed word list from a published corpus. Inflected forms are listed separately (for all languages except Korean) and are associated with token frequencies from their source corpus. No grammatical or function words were included in the dataset. A summary of the contents of the database is presented in Tables 1 and 2.

For each of these languages, at least one diachronically-recent phoneme pair merger has been reported in the literature in an otherwise phonemically similar dialect of the language. Assessing a correlation of functional load with phoneme merger probability requires a comparison between phoneme pairs that have merged (historically, or in some dialect of a language) and those that have not. Because phonemes that merge tend to be phonetically similar (Labov, 1994, ch. 11) we limited the comparison set of non-merged phoneme pairs to pairs which differ in only one phonological feature such as *voice* or *place of articulation*. The comparison set for consonant mergers was further limited to unmerged consonant-phoneme pairs for the language, and likewise the comparison set for vowel mergers was limited to unmerged vowel-pairs.

The dataset consists of 18 such comparison sets of phoneme pairs[1] where each phoneme pair in each comparison set is coded as "Merged" if it has merged in some phonemically similar dialect of that language, and "Unmerged" if not. In total, the dataset contains 56 phoneme pairs that have merged, and 578 that have not.[2]

### 2.2. Predictor variables

There are many possible operationalizations of the notion of functional load (Hockett, 1967; Kaplan, 2011; King,

1967; Silverman, 2010; Surendran & Niyogi, 2006), differing in relation to unit size, role of frequency, word-category and other variables. A particularly simple measure is the number of lexical minimal pairs in a corpus that are distinguished by a phoneme contrast.[3] Martinet (1952) and Hockett (1995) proposed that word frequency should be taken into account as well, and Hockett (1967), Surendran and Niyogi (2003), and Surendran and Niyogi (2006) described a general framework for assessing functional load of phonemic contrasts in terms of system entropy at varying levels of analysis. At the word level, we compare the number of minimal pairs in the corpus defined by a phoneme pair to the change in word-level entropy of the corpus upon merger of the phoneme pair (Surendran & Niyogi, 2006). Similarly, at the phoneme level[4] we compare phoneme probability in a corpus to change in the phoneme-level entropy upon merger of the phoneme-pair (Hockett, 1967). Individual phoneme probabilities ($p(\phi_i|C)$) are calculated by the formula in Eq. (1), where $\phi_i$ is a phoneme, $f(\phi_i)$ is its frequency-count in the corpus, assessed at either the word-token or -type level, $C$ is the relevant corpus, and $\Phi$ is the set of all phonemes in that corpus. We tested the predictive value of the higher-probability member of the phoneme pair separately from the lower-probability member of the pair, as well as the sum of their probabilities (Eq. (2)), and the ratio of the probability of the lower-probability phoneme of the pair to the higher (Eq. (3)). Further, we tested these phoneme probability factors with phoneme counts per word type and word token in each corpus. The results we report here use the natural logarithm of these probabilities in the model-fitting analyses, but the untransformed measures provide the same basic results. As discussed below, the formulations of phoneme probability based on the higher-probability phoneme of the pair were superior to the others, and the model reported below uses the natural logarithm of the token-based probability of the higher-probability member of the phoneme pair.

$$p(\phi_i|C) = \frac{f(\phi_i)}{\sum_{\phi \in \Phi} f(\phi)} \tag{1}$$

$$p(\phi_{\text{low}}|C) + p(\phi_{\text{high}}|C) = \frac{f(\phi_{\text{low}})}{\sum_{\phi \in \Phi} f(\phi)} + \frac{f(\phi_{\text{high}})}{\sum_{\phi \in \Phi} f(\phi)} \tag{2}$$

$$\frac{p(\phi_{\text{low}}|C)}{p(\phi_{\text{high}}|C)} = \frac{f(\phi_{\text{low}})}{\sum_{\phi \in \Phi} f(\phi)} \times \left( \frac{f(\phi_{\text{high}})}{\sum_{\phi \in \Phi} f(\phi)} \right)^{-1} \tag{3}$$

Hockett (1967) and Surendran and Niyogi (2006) described a general framework for assessing functional load of phonemic contrasts in terms of the change in system entropy at any level of analysis upon loss of a phoneme contrast. Based on information theoretic methods introduced by Shannon (1948), this approach assumes that a language can be described as an infinite sequence of phonemes, and

---

[1] Because the American and British-based English phoneme inventories are distinct they each contribute independent comparison sets.

[2] Further details concerning the composition of the dataset are provided in an Online appendix.

[3] Minimal pair counts, as well as other factor values, are reported in an Online appendix to this paper. Here we define a minimal pair as any pair of word-forms that are distinguished by a single phoneme contrast. By this definition, both [pɪt ∼ bɪt] and [papa ∼ baba] are examples of minimal pairs defined by the [p ∼ b] contrast.

[4] For the purposes of this analysis, we treat tone in the Hong Kong Cantonese data as a segment-level property of a syllable, rather than, for example, as a feature of a vowel.

**Table 1**
Summary of the database constructed for this study: languages included, phoneme pairs included, and attested mergers.

| Language | Corpus | Phoneme types/context | Actual mergers | Notes | References |
|---|---|---|---|---|---|
| English (*RP*) | CELEX | $V \sim V$ | $aɪ \sim ɔɪ$ | PRICE $\sim$ CHOICE | Wells (1982) |
| | | | $ʊə \sim ɔː$ | CURE $\sim$ THOUGHT | |
| | | | $ɪə \sim ɛə$ | NEAR $\sim$ SQUARE | |
| | | | $ɛː \sim ɛə$ | NURSE $\sim$ SQUARE | |
| | | $C \sim C$ | $θ \sim t$ | | |
| | | | $θ \sim f$ | | |
| | | | $θ \sim s$ | | |
| English (*American*) | CMU pronouncing dictionary | $V \sim V$ | $a \sim ɔ$ | LOT $\sim$ THOUGHT | Labov et al. (2006) |
| | | | $ɔɪ \sim ɛɹ$ | | |
| | | $V \sim V/\_ɹ$ | $a \sim ɔ$ | START $\sim$ NORTH | |
| | | $V \sim V/\_n$ | $ɪ \sim ɛ$ | PIN $\sim$ PEN | |
| | | $V \sim V/\_l$ | $ɪ \sim i$ | HILL $\sim$ HEEL | |
| | | | $ʊ \sim u$ | PULL $\sim$ POOL | |
| | | | $ʊ \sim oʊ$ | BULL $\sim$ BOWL | |
| | | | $ʌ \sim a$ | HULL $\sim$ HALL | |
| | | | $ʊ \sim ʌ$ | BULL $\sim$ HULL | |
| | | $C \sim C$ | $ʍ \sim w$ | | |
| | | $C \sim C/\acute{V}\_V$ | $t \sim d$ | flapping | |
| German | CELEX | $V \sim V$ | $e \sim ɛː$ | GEBE $\sim$ GÄBE | Wiese (2000) |
| Dutch | CELEX | $C \sim C$ | $s \sim z$ | | Kissine et al. (2003) |
| | | | $f \sim v$ | | |
| | | | $x \sim ɣ$ | | |
| French | Lexique | $V \sim V$ | $\tilde{ɛ} \sim \tilde{œ}$ | VIN $\sim$ UN | Fagyal et al. (2006) |
| | | | $e \sim ɛ$ | ÉPÉE $\sim$ ÉPAIS | |
| | | | $œ \sim ø \sim ɔ$ | | |
| Spanish | Spanish Gigaword | $C \sim C$ | $ʎ \sim j$ | | Penny (2002)[106] |
| | | | $s \sim θ$ | | Harris (1969) |
| Slovak | Slovak National Corpus | $C \sim C$ | $ʎ \sim l$ | | Krajčovič (1988) |
| | | $V \sim V$ | $æ \sim a$ | | |
| | | | $æ \sim e$ | | |
| Korean | Korean Academy Database | $C \sim C/[-son]\_$ | $p \sim p'$ | post-obs. tensing | Sohn (1999)[173] |
| | | | $t \sim t'$ | | |
| | | | $s \sim s'$ | | |
| | | | $ʧ \sim ʧ'$ | | |
| | | | $k \sim k'$ | | |
| | | $C \sim C/\_]_σ$ | $p \sim p^h$ | coda neut. | Sohn (1999)[165] |
| | | | $t \sim t^h \sim s \sim s'$ | | |
| | | | $\sim ʧ \sim ʧ^h \sim h$ | | |
| | | | $k \sim k^h \sim k'$ | | |
| Cantonese (*Hong Kong*) | Hong Kong Cantonese Corpus | $C \sim C/\#\_$ | $n \sim l$ | | Zee (1999) |
| | | $C \sim C/\_\#$ | $n \sim ŋ$ | | Zee (1985) |
| | | $T \sim T$ | $2 \sim 5$ | | Mok and Wong (2010) |

that a corpus is a sample of this sequence. Following Surendran and Niyogi (2006), we calculated the change in entropy of the corpus upon phoneme merger at both the phoneme and the word level. The entropy *H* of the corpus *C* can be used as an estimate of the entropy of the language, as in Eq. (4), where *x* is an element from the set *X*, and $p(x)$ is the measured probability of the element *x* in *C*. To calculate the entropy at the phoneme level, *X* is simply the set of phoneme types in the corpus, while for example to calculate the entropy at the word level, *X* is the set of word types in the corpus.

$$H(C_x) = -\sum_{x \in X} p(x) \log_2 p(x) \qquad (4)$$

The change in entropy of the corpus upon loss of a phoneme contrast $a \sim b$, $\Delta H(C_{X,a \sim b})$, is given in Eq. (5), where $H(C_x)$ is the entropy of the corpus measured at some level *X*, while $H(C_{X,a \sim b})$ is the entropy calculated with all phonemes *a* and *b* merged by replacement with a single symbol.

$$\Delta H(C_{X,a \sim b}) = H(C_X) - \frac{H(C_{X,a \sim b})}{H(C_X)} \qquad (5)$$

When measured at the phoneme level, phoneme merger necessarily results in a change in system entropy because each phoneme has some probability in the corpus. However, when measured at some higher level, such as the word, phoneme merger results in a change in system entropy only if that phoneme contrast distinguishes elements at that level. As an example, if a phoneme contrast distinguishes no minimal word pairs in the corpus, its merger will not result in a change in system entropy at the word level.

We compare the minimal pair and phoneme probability measures to the word- and phoneme-level entropy

**Table 2**
Summary of the corpora from which data were taken.

| Language | Corpus | Corpus size | | Data sources |
|---|---|---|---|---|
| | | Lemmas | Wordforms | |
| English (RP) | CELEX (Baayen et al., 1995) | 52,447 | 160,595 | Words from dictionaries; frequencies from written American English texts |
| English (American) | CMU pronouncing dictionary[a] (Weide, 1995) | NA | 133,354 | Dictionaries; dictionaries of proper names |
| German | CELEX (Baayen et al., 1995) | 51,728 | 365,530 | Frequencies from newspapers, fiction and non-fiction books, and transcribed speech |
| Dutch | CELEX (Baayen et al., 1995) | 124,136 | 381,292 | Dictionaries; fiction and non-fiction books |
| French | Lexique (New et al., 2001) | 53,082 | 142,694 | Novels; film subtitles |
| Spanish | Spanish Gigaword 3rd Edition (Mendonça et al., 2009) | NA | 41,633[b] | Newswire text data |
| Slovak | Slovak National Corpus (Šimková, 2006) | NA | 172,470 | Newspapers; books; other written texts |
| Korean | Korean Academy Database (Lee, 2006) | 58,437 | NA | |
| Cantonese (Hong Kong) | Hong Kong Cantonese Corpus (Leung & Law, 2009) | NA | 170,000 | Phone-in radio programs |

[a] The CMU pronouncing dictionary lacks frequency information. Frequencies for this dataset were taken from CELEX.
[b] Content words only, identified automatically by TreeTagger (Schmid, 1995).

measures, respectively, because while they are comparable in their levels of analysis they differ in the degree to which properties of the system as a whole are taken into account. The minimal pair and phoneme probability measures are *local* in the sense that they do not depend on the number or probability of other word or phoneme types in the system. The entropy measures do take these relationships in the rest of the system into account, and as such the two types of measures imply different models for mechanisms underlying an effect of functional load on phoneme merger. While a thorough analysis comparing the predictive power of other functional load measures is beyond the scope of this short report, the general pattern of results reported here holds for most of the formulations we have investigated.

### 2.3. Results

In this section, we report a hierarchical (or *mixed-effects*) logistic regression model (Baayen, 2008; Gelman & Hill, 2007; Jaeger, 2008). The grouping factor (i.e., *random effect*) in the hierarchical model is phoneme comparison set, defined above in Section 2.1. It is possible to treat this factor as nested within levels of language/corpus. However, neither this nor other reasonable random effect structures, including random slope effects, improved on a basic model with a random intercept for phoneme comparison set as assessed by model selection based on AIC, BIC, and $\chi^2$ likelihood ratio tests. We therefore report the fixed effects only from this simpler model. The use of a hierarchical model is preferred for this kind of structured data, and the fact that this model represents "partial pooling" of results across comparison sets and 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.

The primary predictor variables (minimal pair counts and phoneme probabilities) were centered and standardized prior to model-fitting. The fixed effects in the final model are presented in Table 3, but several steps were taken to arrive at this model. First, we fit a simpler model with no interaction between minimal pair counts and phoneme probabilities. In this model, both predictors are significant, in opposite directions: the more minimal pairs distinguished by the phonemes, the less likely merger is ($\beta = -3.55$, $z = -3.93$, $p < 0.001$), but the more probable the phoneme, the *more* likely merger is ($\beta = 0.71$, $z = 3.65$, $p < 0.001$).

In exploring this model, we found that a straightforward linear interaction between these two variables proved non-significant ($p = 0.462$), but closer inspection revealed that there was an abrupt shift in the effect of phoneme probability between the pairs distinguishing no word minimal pairs vs. those with minimal pairs. In the model presented here, this is captured by including a dichotomous predictor with the levels 'minimal pairs' and 'absence of minimal pairs,' which is allowed to interact with phoneme probability.[5] This complication to the model is justified by both a $\chi^2$ likelihood ratio test ($\chi^2 = 7.42$, $df = 2$, $p = 0.025$) and the fact that the interaction is significant in the model (as seen in Table 3).

We also established that the probability of the higher-probability member of the phoneme pair was a better predictor than alternatives such as the probability of the lower-probability member, the summed probability of the pair, or their ratio. This was tested by fitting superset models which included both the higher-probability predictor and the competing predictor, and comparing those models to subset models that included either the higher-probability predictor or the competitor. In each case, $\chi^2$ likelihood ratio tests confirmed that the superset model did not predict better than the model in Table 3 (all $p$'s > 0.20), but did predict better when the higher-probability phoneme probability variable was replaced by a competitor (for

---

[5] This categorical variable was modeled using simple treatment coding, with 'presence of minimal pairs' as the reference (intercept) level.

**Table 3**
Fixed effects in logistic mixed-effects model. Standard deviation of random intercept = 0.89.

| Predictors | Estimate | Std. error | z Value | p Value |
|---|---|---|---|---|
| (Intercept) | −3.33 | 0.43 | −7.82 | <0.001 |
| Minimal pairs | −3.34 | 0.92 | −3.65 | <0.001 |
| Phoneme probability | 0.40 | 0.23 | 1.77 | 0.076 |
| Absence of min. pairs (binary) | −0.51 | 0.50 | −1.02 | 0.306 |
| Phoneme freq. by absence of min. pairs | 1.53 | 0.72 | 2.11 | 0.035 |
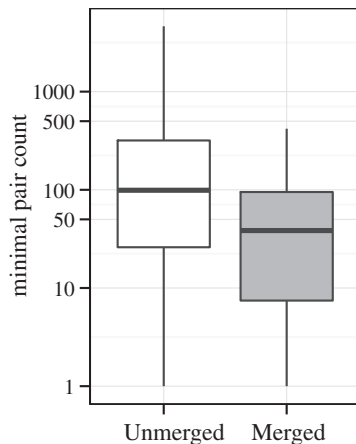


Fig. 1. Relationship between minimal pair count and merger.



Fig. 2. Relationship between phoneme probability and merger, by presence of minimal pairs.

the lower-probability variable: $\chi^2$ = 23.13, $df$ = 2, $p$ < 0.001; for the summed probability variable: $\chi^2$ = 7.19, $df$ = 2, $p$ = 0.027; for the probability ratio: $\chi^2$ = 8.09, $df$ = 2, $p$ = 0.017). These results provide evidence that the probability of the higher-probability phoneme in a pair is the more important predictor in this data.

Finally, the predictive value of phoneme probabilities based on word-type or -token counts were not significantly different, and here we present the variable based on the word-token count. The parameters of this final model are the ones given in Table 3. The model supports the basic functional load hypothesis: the negative effect of minimal pairs illustrated in Fig. 1 is consistent with the claim that when a phoneme plays a greater role in distinguishing words, it is more resistant to merger processes. The magnitude of this effect can be interpreted from the coefficients. Because the predictors are standardized, the coefficients in Table 3 are interpreted as the change in the log-odds of merger for a change of one standard deviation in the predictor. In this data set, the mean number of minimal pairs was 218 and the standard deviation was 446. For example, the model says that all else being equal, the probability of merger with 10 minimal pairs is 0.146 (log-odds -1.77), so the probability of merger with (10 + 446=) 456 minimal pairs is 0.006 (log-odds -5.11).

The effect of phoneme probability is more subtle because of its interaction with the presence/absence of minimal pairs. In the model in Table 3, the presence of minimal pairs is taken as the 'baseline', and the borderline-significant simple effect of phoneme probability is interpreted to mean that for phoneme pairs which distinguish minimal pairs in the language, phoneme probability does not play a
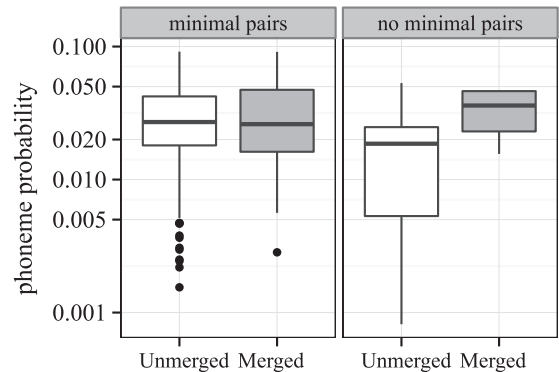
reliable role in predicting merger. However, the significant interaction indicates that where there are no minimal pairs, merger is more likely for higher-probability phonemes. Fig. 2 shows this graphically.

We note that as expected, phoneme probability and number of minimal pairs are positively correlated ($r$ = 0.31, $t$(632) = 8.28, $p$ < 0.001), as shown by the scatterplot in Fig. 3.[6] In order to rule out the possibility that the model was adversely affected by this collinearity, we residualized the phoneme probability variable on the minimal pair variable and refit the model. The results were virtually identical to those in Table 3, suggesting that collinearity is not a major concern in interpreting the model results.[7]

Having established the model in Table 3, we ask whether entropy-based measures improve the model. We consider two measures of entropy, word-level entropy and phoneme-level entropy, as described in Section 2.2. Because word-level entropy is more correlated with our minimal pairs variable than the phoneme probability variable (0.72 vs. 0.32, respectively), we considered whether word-level entropy could be a more effective replacement

---

[6] For clarity, this figure excludes observations with extreme values of minimal pair counts.

[7] The pattern in Fig. 3 does suggest a reason why the effect of phoneme probability is diminished in the presence of minimal pairs. Above zero minimal pairs, there is an abrupt narrowing of the distribution of phoneme probability to values above around -5 log phoneme probability. In our data, the interaction effect between phoneme probability and the dichotomous minimal pair factor becomes marginal when excluding data below -5 log phoneme probability. Consequently, it is possible that the interaction between minimal pair existence and phoneme probability is a simple by-product of the lack of very low phoneme probabilities given the existence of minimal pairs.
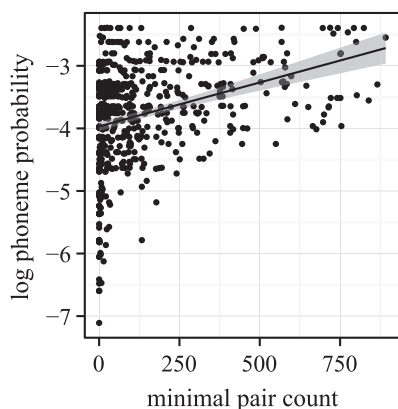
**Fig. 3.** Relationship between minimal pair count and (log) phoneme probability. Shaded area represents 95% confidence interval around the regression line slope.

predictor for minimal pair count. Conversely, phoneme-level entropy is more correlated with phoneme probability in our data (0.56 vs. 0.6 with minimal pairs), so we considered these two predictors as competitors.[8] To assess the relative effectiveness, we first fit a superset model including both competitors. We then fit an alternative model substituting the entropy variable in place of the competitor in our model in Table 3. Using $\chi^2$ likelihood ratio tests, we compared both our model in Table 3 and the competing entropy-based model to the superset model.

In the case of word-level entropy vs. minimal pairs, model comparison indicated that the superset model was significantly better-fitting than the word-entropy model ($\chi^2 = 9.61$, $df = 1$, $p = 0.002$), but not significantly better than the minimal pairs model without word entropy ($\chi^2 = 0.4$, $df = 1$, $p = 0.527$). Similarly, the superset model including both phoneme probability and phoneme level entropy was significantly better-fitting than the phoneme-entropy model ($\chi^2 = 23.7$, $df = 2$, $p < 0.001$), but not better than the phoneme probability model without phoneme entropy ($\chi^2 = 3.41$, $df = 2$, $p = 0.182$). In summary, entropy-based formulations of functional load at the word level or phoneme level did not improve on the model presented in Table 3.

## 3. Conclusions

This paper reports the first statistical evidence that the probability of phoneme merger is indeed inversely related to functional load, as predicted for nearly a century. We accomplished this by employing a statistical analysis of a relatively large dataset drawing on a variety of languages, rather than on individual case studies. Within this dataset, we find that the more minimal pairs defined by a phoneme pair, the less likely that phoneme pair is to have merged. These results provide the first clear support for the general intuition behind the functional load hypothesis, which is

that merger is less likely between phonemes which contribute more to distinguishing words.

This finding is consistent with models that propose a causal chain linking individual utterances to long-term change in the abstract, sublexical category system of a speech community (Beckner et al., 2009; Blevins, 2004; Blevins & Wedel, 2009; Bybee, 2001; Hay & Maclagan, 2012; Kirby, 1999; Labov, 1994; Ohala, 1989; Pierrehumbert, 2001; Pierrehumbert, 2003; Walsh, Möbius, Wade, & Schütze, 2010; Wedel, 2007). More specifically, Wedel (2004), Wedel (2006), and Wedel (2012), describes a multi-level exemplar model (Walsh et al., 2010) of a general linking mechanism between biased variation in word production/perception events and long-term change in phonetic distributions within sound and word categories. In this model, any mechanism in production or perception favoring phonetically more contrastive tokens of minimal pair members promotes greater phonetic distinction between the phonemes defining that minimal pair across the lexicon. A prediction of this model is that the distinguishing phonemes in tokens of minimal pairs will be produced, on average, with more exaggerated phonetic cues. Consistent with this prediction, a number of studies have shown that phonetic cues to linguistic categories such as syllables and words are enhanced when the uncertainty of that category is greater in context (Aylett & Turk, 2004; Cohen Priva, 2012; Jurafsky, Bell, Gregory, & Raymond, 2001; Kaiser, Li, & Holsinger, 2011; Piantadosi, Tily, & Gibson, 2011; Raymond, Dautricourt, & Hume, 2006; van Son & Pols, 2003). Similarly, Baese and Goldrick (2009) and Peramunage, Blumstein, Myers, Goldrick, and Baese-Berk (2011) report that when read aloud from a list, initial stops in English words that have a minimal pair defined by the voicing of the initial stop (e.g., bat ~ pat) are produced with stronger phonetic cues to stop-voicing. Peramunage et al. (2011) further report that enhancement in the VOT cue to stop-voicing was not significantly predicted by lexical neighborhood density[9] as has been proposed to account for similar findings in read-speech (Munson, 2007; Munson & Solomon, 2004; Scarborough, 2010; Wright, 2004), but see (Gahl, Yao, & Johnson, 2012). In support of a specific minimal pair effect on production of phonetic cues, rather than a more general effect of lexical neighborhood density, we tested a variety of neighborhood density factors (see, e.g., Scarborough, 2010) in the model and found no significant predictive effect on merger probability.

When there are few or no minimal pairs distinguished by a phoneme pair, a second finding is that greater phoneme probability, at least for the higher-probability member of the pair, is significantly associated with merger. This is consistent with linguistic evidence (Bybee, 2001; Hay & Maclagan, 2012), and theoretical models (Bybee, 2001; Hay & Maclagan, 2012; Walsh et al., 2010) suggesting that the rate of change in the distribution of cues to lexical and sublexical categories is correlated with frequency of use. We did not find a significant effect of the ratio of probabilities of the members of the phoneme pair as suggested by

---

[8] Treating phoneme-level entropy as a competitor to minimal pairs instead produced the same pattern of results as in Table 3, but with a worse AIC, suggesting that it fares no better as a replacement for minimal pairs.

[9] Neighborhood density measures are based on the number of lexical items within a string edit-distance of one with respect to the reference item.

models which propose a higher error bias between phoneme categories of divergent frequencies (Pierrehumbert, 2001; Wedel, 2006). However, the relatively small size of the dataset makes it difficult to draw strong conclusions in this regard.

Hockett (1967), followed by Surendran and Niyogi (2003, 2006), note that the change in system entropy upon loss of a contrast is the most direct index of that contrast's contribution to the overall information transmission capacity of the system (Shannon, 1948) and argue that it is, by extension, the appropriate measure of functional load. In this dataset, however, we find that the more local measures of minimal pair count and relative phoneme probability do a significantly better job of predicting merger. There are two possible reasons for this finding. One is simply that the properties of the corpora on which the dataset is based, in particular their frequency information, may not correspond sufficiently well to the language context during the merger process. The other is that a putative contrast-maintenance mechanism may not directly take the contrast properties of the entire system into account (as is implicit in the use of an entropy-based measure), but instead may operate on the basis of, for example, a more local interaction between individual word (Baese & Goldrick, 2009) and phoneme categories (Pierrehumbert, 2001). We note, however, that any mechanism maintaining contrast at these more local levels will tend to support the information-transmission capacity of the system as a whole.

## Acknowledgement

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2013.03.002.

## 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. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R.H., Piepenbrock, R., & van Rijn, H. (1995). *The CELEX lexical database. 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. (2004). *Evolutionary phonology*. Cambridge: Cambridge University Press.

Blevins, J., & Wedel, A. (2009). Inhibited sound change: An evolutionary approach to lexical competition. *Diachronica, 26*, 143–183.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Cohen Priva, U. (2012). *Sign and signal: Deriving linguistic generalizations from information utility*. Unpublished doctoral dissertation, Standford University.

Fagyal, Z., Kibbee, D., & Jenkins, F. (2006). *French: A linguistic introduction*. Cambridge: Cambridge University Press.

Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language, 66*, 789–806.

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'apres l'Atlas Linguistique de la France*. Paris: É. Champion.

Harris, J. (1969). *Spanish phonology*. Cambridge, MA: MIT Press.

Hay, J., & Maclagan, M. (2012). /r/-Sandhi in early 20th century New Zealand English. *Linguistics, 50*, 745–763.

Hockett, C. (1960). The origin of speech. *Scientific American, 203*, 89–96.

Hockett, C. (1967). The quantification of functional load. *Word, 23*, 320–339.

Hockett, C. (1955). *A manual of phonology*, Bloomington. IN: Indiana University.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jakobson, R. (1931). Prinzipien der historischen Phonologie. *Travaux du cercle linguistique de Prague, 4*, 246–267.

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* (pp. 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). Cambridge, MA: Springer. 171–183.

Kaplan, A. (2011). How much homophony is normal? *Journal of linguistics, 47*, 631–671.

King, R. D. (1967). Functional load and sound change. *Language, 43*, 831–852.

Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford: Oxford University Press.

Kissine, M., van de Velde, H., & van Hout, R. (2003). An acoustic study of standard Dutch /f/, /v/, /s/, and /z/. In L. Cornips & P. Fikkert (Eds.), *Linguistics in the Netherlands* (pp. 93–104). Amsterdam: John Benjamins.

Krajčovič, R. (1988). *Vývin Slovenského jazyka a dialektológia*. Bratislava, Slovakia: Slovenské Pedagogické Naklatel'stvo.

Labov, W. (1994). *Principles of linguistic change. Internal Factors* (Vol. I). Oxford: Blackwell.

Labov, W. (2001). *Principles of linguistic change. Social Factors* (Vol. II). 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.

Leung, M.-T., & Law, S.-P. (2009). HKCSC: The Hong Kong Cantonese adult language corpus. *International Journal of Corpus Linguistics, 6*(2), 305–325.

Martinet, A. (1952). Function, structure, and sound change. *Word, 8*, 1–32.

Mathesius, V. (1931). Zum Problem der Belastungs- und kombinationsfähigkeit der Phoneme. *Travaux du cercle linguistique de Prague, 4*, 148–152.

Mendonça, A., Graff, D., & DiPersio, D. (2009). *Spanish gigaword second edition*. Philadelphia, PA: Linguistic Data Consortium.

Mok, P.M.-K., & Wong, P.W.-Y. (2010). Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. In *Fifth international conference on speech prosody 2010*. Chicago: International Speech Communication Association.

Munson, B. (2007). Lexical access, lexical representation, and vowel production. In J. S. Cole & J. I. Hualde (Eds.). *Papers in laboratory phonology* (Vol. IX, pp. 201–222). New York: Mouton.

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'anné Psychologique, 101*, 447–462.

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.

Penny, R. J. (2002). *A history of the Spanish language* (2nd ed.). Cambridge: Cambridge University Press.

Peramunage, D., Blumstein, S. E., Myers, E., Goldrick, M., & Baese-Berk, M. (2011). Phonological neighborhood effects in spoken word production: An FMRI study. *Journal of Cognitive Neuroscience, 23*, 593–603.

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* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115–154.

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.

Scarborough, R. (2010). Lexical and conceptual predictability: Confluent effects on the production of vowels. In C. Fougeron, B. Kühnert, M. D'Imperio, & N. Vallée (Eds.). *Papers in laboratory phonology* (Vol. 10, pp. 557–586). Berlin: de Gruyter.

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-workshop*. Dublin: Association for Computational Linguistics.

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.

Šimková, M., (2006). Slovak National Corpus – History and current situation. In M. Šimková (Ed.), *Insight into the Slovak and Czech corpus linguistics* (pp. 151–159). Bratislava, Slovakia: Veda.

Sohn, H.-H. (1999). *The Korean language*. Cambridge: Cambridge University Press.

Studdert-Kennedy, M., & Goldstein, L. (2003). Launching language: The gestural origin of discrete infinity. In M. H. Christiansen & S. Kirby (Eds.), *Language evolution: The states of the art*. Oxford: Oxford University Press.

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* (pp. 43–58). Amsterdam: John Benjamins.

Trubetzkoy, N. (1939). *Grundzüge der phonologie*. Prague, Czech Republic: Travaux du Cercle Linguistique de Prague.

van Son, R.J.J.H., & Pols, L.C.W. (2003). How efficient is speech? In E.H. Berkman (Ed.), *Proceedings of the institute of phonetic sciences*. Amsterdam.

Walsh, M., Möbius, B., Wade, T., & Schütze, H. (2010). Multilevel exemplar theory. *Cognitive Science, 34*, 537–582.

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. (2012). Lexical contrast maintenance and the development of sublexical contrast systems. *Language and Cognition, 4*, 319–355.

Weide, R.L. (1995). *The CMU pronouncing dictionary. Carnegie Mellon University* <Retrieved http://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, pp. 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.