Andrew Wedel

Lexical contrast maintenance and the organization of sublexical contrast systems

Abstract: Variationist/evolutionary models of phonology assume a causal chain that links biases at the utterance level to the development and consolidation of abstract phonological patterns over time. Some of the properties of linguistic cognition that have been proposed to underlie this chain are (i) storage of experienced detail at multiple levels of description, (ii) feedback between perception and production, (iii) a similarity bias in the production and perception of variation, and (iv) enhancement of cues to potentially ambiguous lexical items in usage. I review evidence for these properties and argue that they interact to provide a pathway for individual usage events to influence the evolution of contrastive sublexcal category systems, i.e phoneme inventories. Specifically, the proposed Network-Feedback model predicts that the organization of sublexical category systems is shaped by a conflict between a general drive toward greater similarity among sublexical categories on the one hand, and a bias toward maintaining contrast between tokens of competing lexical categories on the other. The model provides testable hypotheses about the conditions favoring phoneme merger, chain-shifts, and phonemic splits, and more generally about the influence of lexical contrast on the packing of sublexical categories along gestural/perceptual dimensions. Finally, this pathway of change is consistent with proposals that sublexical categories such as features and segments are not primitives of language, but emerge through more general properties of performance, perception, categorization and learning.

Keywords: functional load, contrast, phoneme, inventory, simulation, phonology, language change, duality of patterning, self-organization

Andrew Wedel: Department of Linguistics, University of Arizona, Tucson, AZ, 85721-0028.
E-mail: wedel@email.arizona.edu

1 Introduction

On both theoretical and evidential grounds, many linguists have argued that patterns of change within sublexical category[1] systems may be influenced by the functional role that sublexical categories play in distinguishing lexical meanings in communication (Gilliéron 1918; Trubetzkoy 1939; Martinet 1952; Hockett 1967; Surendran and Niyogi 2006; Silverman 2010; Kaplan 2011; Blevins and Wedel 2009; Wedel et al. In press). At the same time, an independent strand of research (e.g. Lindblom et al. 1984; Joanisse and Seidenberg 1997; de Boer 2001) has proposed that pressure to maintain sublexical category distinctions contributes to the observed trend toward symmetry and economical packing of sublexical categories along articulatory and perceptual dimensions (Maddieson 1984). In this paper I argue that a set of properties present in the general variationist/usage-based/evolutionary model of sound change[2] interact to provide a causal link between support of lexical contrast in individual usage events and the long-term evolution of sublexical category inventories. Because this 'Network-Feedback' model predicts that measurable properties of lexical contrast in usage influence sublexical patterns, it provides testable hypotheses about the conditions favoring sound changes such as chain-shifts, mergers, and splits, and about how lexical contrast influences the packing of sublexical categories along gestural/perceptual dimensions. Finally, this pathway is consistent with proposals that sublexical categories such as features and segments are not primitives of language, but emerge and are propagated through more general properties of performance, perception, categorization and learning (e.g. Hockett 1960; Lindblom et al. 1984; Kirby 1999; Oudeyer 2002; Smith et al. 2003; Studdert-Kennedy and Goldstein 2003; Mielke 2008; see also Sandler et al. 2011). In the following section, I review evidence for four general properties of linguistic information storage and transmission that form the basis for the model:

---

1 As this paper concerns the gradient development and loss of sublexical cues to lexical contrast, I will avoid using the term *phoneme* except when referring to others' work. Instead, I will use the term *sublexical category* to refer to any conventionalized perceptual or gestural unit below the lexical level.

2 I use the cover term 'variationist/usage-based/evolutionary' for the many theoretical and experimental contributions based in psycho- and sociolinguistics, cognitive science, evolutionary theory, and phonetics that provide new hypotheses about how variation at multiple levels and time-scales can drive the development, propagation and consolidation of phonological patterns over time, (e.g. Lindblom et al. 1984; Ohala 1989; Labov 1994, 2001; Kirby 1999; Bybee 2001; Hume and Johnson 2001; Pierrehumbert 2002; Blevins 2004; Wedel 2007; Mielke 2008; Beddor 2009, and many more).

- Storage of experienced detail at multiple levels of description
- Feedback between perception and production
- A similarity bias in produced and perceived variation[3]
- Enhancement of cues to the identity of potentially ambiguous lexical items

The first three properties together allow production and perception biases, including a general bias toward similarity, to initiate and extend patterns within and between categorial levels in the lexicon. In earlier work, I have argued that the tendency of speakers to reproduce perceived phonetic variation, in combination with their tendency to extend patterns, interacts with the effect of small biases in individual language usage events to influence the generation and spread of phonological patterns through a speech community (Wedel 2004a, 2007, 2009). In Section 2, I review evidence for these properties, including evidence for a usage-based bias toward hyperarticulation of potentially ambiguous lexical items. In Section 3, I explain how these properties, when iterated over many usage events and across many speakers, provide an account for the development and maintenance of a constrained system of sublexical category contrasts, that is to say, a lexicon characterized by duality of patterning (Hockett 1960; Martinet 1960). Because causal mechanisms that integrate many small and local interactions over time are not easily visualized, in Section 4, I introduce a model computational system that exhibits these four general properties (see Wedel 2011 for a primer on *self-organization* as a general structure-formation mechanism). This system illustrates how these properties interact to produce an analogue of a lexical system exhibiting duality of patterning, and how a bias toward maintenance of contrast at a higher level of organization can drive the formation of coherent patterns at a lower level.

2 Components of the network feedback model

2.1 Levels of organization in memory

A common model of the phonological lexicon builds hierarchical structure progressively through a set of nested abstract categories, where each category is

---

3 *Similarity* is a cover term used here to describe any sets of properties that promote cognitive, perceptual or behavioral association. In phonology, some of these properties arise from physical and cognitive architectures generally shared between humans; others may result from the idiosyncratic properties and category systems of a particular language (discussed in Blevins 2004).

composed from an organized set of category labels at a smaller granularity (see e.g. Kenstowizc 1994 reviewed in Baayen 2007; Ernestus 2011). Words, for example, are defined as organized sets of phoneme labels, which are in turn defined as organized sets of features. In this type of model phonetic detail makes contact with the lexicon only via the lowest categorial level, e.g. the feature. This general model has been successful in part because it can explain the striking degree of regularity exhibited by phonological grammars: if a /t/ is composed of the same features no matter what lexical category it may find itself in, we can explain why the phonetic realization of /t/ is predictable across contexts. If on the other hand, every lexical category were able to freely evolve its own relationship to the phonetic production system, it would not be immediately clear why there should be any phonological regularity at all (see Pierrehumbert 2002 and Wedel 2007 for discussion of the tension between evidence for the storage of phonetic variation and observed phonological pattern regularity).

However, a large body of research suggests that categories at higher levels do maintain some record of experienced variation, rather than consisting solely of abstract generalizations (reviewed in Johnson 1997; Bybee 2002b; Pierrehumbert 2002; Baayen 2007; Pisoni and Levi 2007; Ernestus 2011). At the lexical level, a number of studies have demonstrated word-specific pronunciation generalizations (see e.g. Pierrehumbert 2002; Hay and Maclagan 2012) For example, the length of the penultimate vowel in the word *memory* shows a wide range of variation [mɛməri ~ mɛmri] that is not found in the minimally distinct word *mammary* [mæməri ~ *mæmri]. Word-specific generalizations such as this cannot be easily accommodated within a strict hierarchy of nested abstractions, where only the smallest unit makes contact with phonetic detail. Instead, a more productive metaphor may be of a multi-dimensional network, in which some level of experienced detail can be represented at multiple levels of analysis, and where generalizations of varying degree of abstraction can emerge from and coexist with that that detail (see e.g. Elman 1995; Bybee and McClelland 2005; Beckner et al. 2009; Walsh et al. 2010). Consistent with this view, studies of long-term priming of word categories have shown that people retain fine-grained experienced phonetic detail in long-term memory (e.g. Johnson 1997; Ju and Luce 2006; Walker and Hay 2011).

2.2 Production-perception feedback

These memories become active participants in the categories they are mapped to, because both categorization and production behavior have been shown to be influenced by previously experienced phonetic detail. In perception, a wide range

of experiments have shown that listeners re-tune sublexical category boundaries in response to hearing variant tokens of those categories in lexical context, that this re-tuning can be persistent (Norris et al. 2003; Eisner and McQueen 2005; Kraljic and Samuel 2005a, b; see also Walker and Hay 2011), and that it can occur at the feature as well as the phoneme level (Kraljic and Samuel 2005a).

The act of categorizing particular linguistic tokens has been shown to not only influence future categorization behavior, but also to influence future production from those categories (reviewed in Pierrehumbert 2002, 2003). For example, subjects exposed to pronunciations of a set of words by a particular speaker show their own pronunciations shifted towards that of the speaker's for up to a week after exposure (Goldinger 2000; see also Sancier and Fowler 1997; Harrington et al. 2000; Pardo 2006; Nielsen 2007; Sanchez 2011). The experimentally established linkage between past experience and future perception and production behavior suggests that a feedback loop operates in usage. This feedback should promote entrenchment of speech behavior within a community, as phonetic details that are more often perceived are more often produced and vice versa (Pierrehumbert 2001; Wedel 2006).

2.3 Similarity bias in variation

The observations that (i) categories retain phonetic detail and (ii) that phonetic detail influences further production and perception have an interesting implication for models of category change. Because random noise in production and perception steadily introduces new variants beyond the current range of variants, retention and re-use of variation should create a constant pressure for categories to broaden (Pierrehumbert 2001). Because categories do not inexorably broaden with use, this approach suggests that there must be countervailing processes that constrain category variance. In previous work, I have argued that this role is filled by the general tendency for variation in production and perception to be biased toward local maxima in experience (Wedel 2004a, 2006, 2007). In perception, linguistic examples of this *similarity bias* can be seen in the perceptual magnet effect (Kuhl 1991), in which sounds are perceived as being more canonical than they actually are, and in the Ganong effect, in which lexicality influences the perception of ambiguous sounds (Ganong 1980). Within production, a large body of evidence shows that segmental substitution errors are more likely to occur between similar than dissimilar segments (e.g. Shattuck-Hufnagel and Klatt 1979; Frisch 1996; McMillan and Corley 2010), and that segmental substitution errors are more likely when the result is an existing word, indicating that the similarity bias in phonological errors extends beyond the source and target segment to the

lexical context (e.g. Dell 1986; Wright and Frisch 2002). What all of these observations have in common is that variation is often biased toward local maxima on cognitively available dimensions of similarity[4]. For the arguments made in this paper, the important feature of this similarity bias is that it will have the effect over time of constraining the range of variation within and across levels of organization, promoting greater coherence within the system (Wedel 2006). Importantly, if the lexicon is organized as a densely interconnected network of representations at different levels of organization, a general bias towards similarity should promote the spread of patterns both from word to word (Wang 1969; Bybee 2002a; Phillips 2006) and from sound to sound (Kraljic and Samuel 2005a; Mielke 2008).

We can illustrate the ability of a *local* similarity bias on variation to create increasingly coherent *global* behavior using a simple computational model. We start with a checkerboard field of 1000 squares, where each square is randomly assigned one of two colors (Figure 1a). In each round, every square sets its color to be the same as that of a randomly chosen immediate neighbor. Figure 1b shows the field of squares after 200 of rounds, in which we can see that these purely local similarity-biased interactions have resulted in broader categorical pattern. Figure 1c shows the field after 1000 rounds, in which random application of the local similarity bias on color change has extinguished the minority color altogether. An important thing to note from this system is that all else being equal,
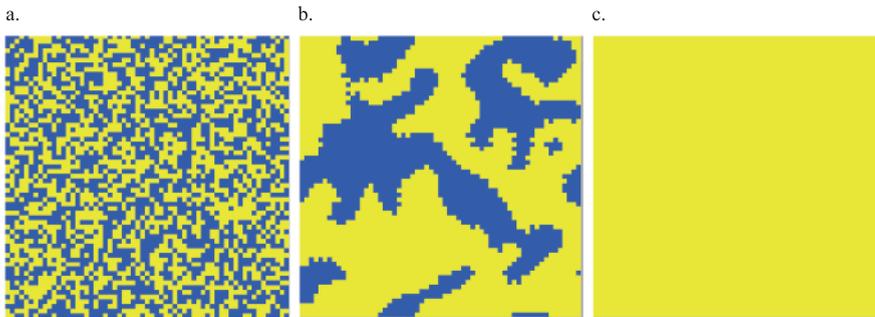
a.   b.   c.

Fig. 1: Large-scale pattern development in a field of elements through similarity-biased local change.

---

4 More generally, research in response biases of cortical fields in both perception and motor behavior suggests that a local similarity bias may arise from general features of neural maps (discussed in Guenther and Gjaja 1996; Oudeyer 2002).

generation of higher-order category distinctions through a local similarity bias is just an intermediate stage on a trajectory toward global uniformity (see Barr 2004 for more complex simulations with abstractly similar outcomes).

2.4 Lexical contrast maintenance

Above, I reviewed experimental evidence for (i) storage of fine-grained, experienced phonetic detail in lexical and sublexical representations, (ii) feedback between perception and production of phonetic detail, and (iii) a similarity-bias in production and perception. In the absence of any mechanism favoring category contrast, these factors together provide a usage-based pathway for categories to merge: given two categories that are adjacent along some dimension of similarity (e.g. voice onset time, or tongue body height), the range of variants produced or perceived from each of two categories will be biased toward the other along that dimension, and storage of these variants will in turn shift the centers of each category along this dimension closer together. If nothing interrupts this feedback loop, with sufficient time the range of variants from these categories will come to overlap along this dimension in the speech of the community (Pierrehumbert 2001; illustrated abstractly in Figure 1 above). If this dimension provides the primary cue distinguishing these two categories, new language acquirers may begin to abstract a single category, at which point the categories have functionally merged (Labov 1994: chapter 11). As it stands then, this general model predicts that category distinctions will tend to merge over the course of time.

And in fact, sublexical category distinctions are often lost over the course of language change (for discussion in this context, see Pierrehumbert 2001). For example, the vowel category /ɔ/ has merged into the category /ɑ/ in Canadian and western American English, such that the originally distinct pronunciation of *caught* ([kɔt]) is now indistinguishable from that of *cot* ([kɑt]; Labov et al. 2006). Collapse of phonetically adjacent categories is not inevitable, however. Instead, we find widespread, although indirect evidence for the existence of some process that inhibits the collapse of sublexical category distinctions over the course of sound change, as in cases of chain shifts (reviewed in Labov 1994: chapter 9; Gordon 2002). In particular, anti-homophony effects within morphological paradigms (e.g. Gessner and Hansson 2004; Blevins and Wedel 2009) suggest that some process inhibiting category collapse may operate at the lexical level.

A number of linguists over the last century have tested the intuitively appealing hypothesis that phoneme contrast loss is inversely related to some measure of the 'work' that the phoneme contrast does in distinguishing lexical items

(Gilliéron 1918; Trubetzkoy 1939; Hockett 1967; King 1967; Surendran and Niyogi 2006; Silverman 2010; Kaplan 2011). However, results of these earlier studies have been equivocal, possibly due to their small sample sizes. Wedel et al. (In press) recently report a statistical analysis of a larger corpus-based dataset comprising 54 phoneme mergers from eight different languages which shows that the number of minimal pairs[5] distinguished by a phoneme contrast is in fact significantly, inversely correlated with phoneme merger. Many sublexical measures are correlated with minimal pair number, such as relative phoneme type or token frequency, but minimal pair number was found to be the strongest predictor of merger independently of all sublexical measures tested. Further, the inhibitory effect on merger was shown to be significantly stronger for minimal pairs that share a syntactic category, e.g. noun~noun, or verb~verb minimal pairs. Minimal pairs that share a syntactic category are less likely to be disambiguated by local morpho-syntactic context and are therefore more potentially ambiguous in usage (Blevins and Wedel 2009). These findings are consistent with the Network Feedback Model described below in which local ambiguity in usage events influences *lexical* representations in such a way as to inhibit *sublexical* category merger over the course of language change.

2.5 Potential mechanisms for lexeme-level contrast maintenance

A diverse body of evidence suggests that in production, cues to lexical category tend to be enhanced when competition from other lexical categories is higher. A number of researchers have reported that in word-reading tasks, vowels in words in high-density neighborhoods tend to be hyperarticulated (Wright 2004; Munson and Solomon 2004; Munson 2007), and that words in high density neighborhoods exhibit increased spreading of cues across neighboring segments (Scarborough 2004). Baese and Goldrick (2009) also report that in a word-reading task, the voice-onset-time (VOT) distinction in initial stop phonemes is exaggerated for words that have a minimal pair defined by initial stop voicing. Two possible sources for these relative hyperarticulation effects are longer-term representations in the lexicon such as neighborhood density and frequency, and shorter-term context-specific influences on lexical predictability. Evidence is consistent with a role for both longer-term, representational factors and local, context-

---

5 For this work, a minimal pair was defined a pair of words that is distinguished by a single phoneme contrast. Under this definition, [bɑb] and [pɑp] are minimal pairs defined by the [b~p] contrast.

related factors. For example, van Son and Pols (2003) report that segments are relatively hyperarticulated in proportion to the disambiguating information they contribute in the sequential parsing of their host word in sentence context. Likewise, Aylett and Turk (2004) report that word frequency, syllable-trigram probability and the givenness (i.e. previous mentions of a word in discourse) are all independently correlated with duration of word-medial syllables. Consistent with this finding, the minimal pair-based VOT hyperarticulation effect reported by Baese and Goldrick (2009) was stronger if both members of a minimal pair were present in the context, but was still significant when only one member was present. Most recently, Scarborough (2010) specifically investigated the relationship of hyperarticulation to lexical neighborhood density and frequency on the one hand and lexical predictability in sentence context on the other. She found that both factors independently influence the duration and dispersion of vowels in a target word.

Contextually and representationally conditioned hyperarticulation effects could arise through a variety of proposed mechanisms, which we can conceptually distinguish by whether the proximal cause of hyperarticulation is located in the speaker or in the listener. In a 'listener-orientation' speaker-based account, speakers may model the state of the listener and bias output targets in order to best aid comprehension, for example by hyperarticulating to make an output easier to correctly perceive (e.g. Lindblom 1990). An alternative online speaker-based mechanism is suggested by Baese and Goldrick (2009), who argue that words with more competitors in production, e.g. words in denser neighborhoods or words with minimal pairs, may be relatively hyperarticulated for the same type of reason such words are more slowly categorized in perception: competition with category neighbors (see Luce and Pisoni 1998). Because lexical items with many competitors (e.g. those in high density neighborhoods) are also those that are more vulnerable to slowed and errorful categorization on the part of the listener (reviewed in Luce and Pisoni 1998), both types of models predict that speakers should more often hyperarticulate the lexical items that are also more difficult for listeners.

Conversely, a listener-based type of model locates the hyperarticulation effect in the listeners of a speech community, who in effect *remove* ambiguous tokens from the category at hand. In support of this possibility, evidence suggests that ambiguous forms may have a higher rate of categorization failure (Luce and Pisoni 1998), and/or may be down-weighted to exert relatively less influence on the distribution of category-token mappings in memory (see e.g. Barr 2004). In addition to any mechanism that directly diminishes the influence of a less contrastive variant on its corresponding category, listener-based errors in categorization have been proposed to indirectly result in contrast maintenance effects

(Wedel 2004b; Blevins and Wedel 2009; see also Labov 1994: 580–588). By any mechanism, if lexically ambiguous speaker variants have a smaller average influence on a listener's future production behavior, perception-production feedback provides a pathway for more contrastive speaker variants to contribute relatively more to the evolving distribution of variants of a lexical item in the speech community.

To summarize, in speaker-based models of lexically-conditioned hyperarticulation, variation is predicted to be biased toward hyperarticulation under conditions of greater lexical competition or discourse ambiguity. In the listener-based model, the range of pronunciation variation associated with lexical categories is less influenced by ambiguous forms, thereby indirectly favoring hyperarticulated forms. Each of these types of models makes testably different predictions, but none yet covers the full range of experimental evidence (reviewed in Baese and Goldrick 2009; cf. Yao 2009). Note however that the mechanisms proposed in these models are not mutually exclusive and additional work may well support a mixed account of lexically conditioned hyperarticulation, including both discourse-specific contrast-supporting mechanisms and influences on longer-term properties of lexical representations (discussed in Wedel 2006; Baese and Goldrick 2009; Scarborough 2010). The two important points for the description of this model here are that there is evidence for lexically specific hyperarticulation of potentially ambiguous utterances, and that there are plausible utterance-level mechanisms that can drive this effect. In the context of production/perception feedback, this contrast-maintaining hyperarticulartion effect should bias the distribution of word pronunciation variants toward sufficient contrast over time in the speech community. I will argue below that although lexically-specific, this support for contrast can maintain not only greater lexical category contrast for specific lexical items, but indirectly to also maintain lexicon-wide sublexical category contrasts, that is, duality of patterning.

3 The network feedback model

We have two observations that seem related, but which reference radically different time-scales: (i) words that have more lexical competitors tend to be relatively hyperarticulated in individual usage events, and (ii) phoneme contrasts that distinguish more minimal pairs are less likely to merge over time (Wedel et al. In press). How can the first process, which appears to be both lexically and contextually specific (van Son and Pols 2003; Aylett and Turk 2004; Baese and Goldrick 2009; Scarborough 2010), create coherent behavior of a segment category across the entire lexicon? For example, even if there is a lexical contrast maintenance

mechanism that promotes a distinction in the VOT of the initial stops of *pat* and *bat*, how can this mechanism influence the VOT distinction in the same initial stops in words without corresponding minimal pairs, such as *passion* and *badger*? The corresponding minimal pair counterparts *bassion* or *padger* do not exist, so under a rich-memory model in which lexical categories can maintain idiosyncratic pronunciation details, we might predict that phonetically similar sublexical categories would drift together and merge wherever that category distinction is unnecessary for lexical contrast. Instead, evidence suggests that although phonemic contrasts may be relatively hypoarticulated when they contribute little to lexical contrast in particular lexical item, they do not inevitably merge in those items. Instead, despite the evidence for word-specific phonetic variation reviewed above, it remains broadly the case across languages that lexical items can be decomposed into subparts that map to a limited inventory of contrastive sublexical categories (see Currie-Hall 2009 for a recent review), even in non-disambiguating positions in a lexical item.

What is needed is a mechanism that can provide a causal link between a contrast-maintenance process operating on individual lexical items in specific usage events and the long-term maintenance of an abstract, lexicon-wide pattern of contrastive sublexical categories. The solution proposed here is based in the body of evidence, reviewed in Section 2.1. above, that variation is recorded at multiple levels of organization, and that variant pronunciations of a given segment or feature in one word can influence the pronunciation of the 'same' segment or feature in another (e.g. Bybee 2002a; Bybee and McClelland 2005; Phillips 2006; Pierrehumbert 2006; Kraljic and Samuel 2005a; Nielsen 2007; Walsh et al. 2010; Hay and Maclagan 2012). This spread and consolidation of variants between and across levels of organization opens up what appears at first to be a counter-intuitive possibility: that a sublexical contrast maintenance effect can be based *interior* to the chain of nested levels of organization, originating at the lexical level rather than at the sublexical level itself. Given consolidation of variants through similarity bias in usage, any process that increases sublexical category contrast in *some* lexical items should indirectly promote contrast for that sublexical category in *all* lexical items that contain it.

Biological evolution provides a close conceptual parallel to this model's account for the ability of a local lexical contrast-maintenance effect to drive global sublexical contrast maintenance throughout the lexicon. A biological population is made up of individuals, each of which contains a set of genes that influence that individual's properties. Over time, errors in DNA replication and repair steadily introduce variation into DNA sequences, with the result that multiple variants of a given gene are often present within a population at any given time. If a particular gene variant provides a relative fitness advantage to the individuals

that possess it, that variant is likely to spread over time through the population at the expense of other variants of that gene. However, it is often the case that a gene variant confers a fitness advantage only in specific contexts, for example when paired with a particular variant of another gene, or in a particular environment. Nonetheless, if a variant confers a fitness advantage on some individuals some of the time, this can still be sufficient to drive its spread through the population over time – even though in most contexts the variant confers no fitness advantage at all. Similarly, even if a lexical contrast-maintenance mechanism promotes hyperarticulation of tokens of a sublexical category in only a subset of words, spread and consolidation of patterns between words may promote the persistence of that sublexical category as a coherent entity across the entire lexicon (Wedel 2004b).

4 Computational illustrations

In this section, I illustrate the workings of the Network Feedback Model using a simple computational system (de Boer 2006). This system does not represent a simulation of language in any direct way, but should instead be understood as an object in its own right which transforms information via the same kinds of interlocking pathways proposed above for language. For example, this computational system is implemented in terms of exemplars, but this is not intended as a claim about the specific nature of mental representations (see e.g. Hare and Elman 1995 for a distributed, connectionist architecture that accomplishes similar multi-level storage and spread of variation). Rather, an exemplar architecture is a particularly transparent way to computationally implement the observation that phonetic variation can be stored at multiple levels of organization and can spread between and across those levels. By implementing these pathways as transparently as possible, and then observing this system's behavior, we can better see how these pathways interact to produce changes in the system that are parallel to well-known patterns of linguistic change relating to sublexical category contrast. In addition, we can view these kinds of computational systems as tools for generating hypotheses that can be tested through other means, such as psycholinguistic or corpus studies. In the following sections, after outlining the structure of the computational system, I show how these pathways of information flow interact to create system behaviors that are parallel to:

– Formation of a contrastive, sublexical category system shared across the lexicon, i.e. a phoneme inventory
– Maintenance of contrast of sublexical categories via lexical contrast maintenance

- Lexical diffusion (Phillips 2006)
- Chain-shifts (Labov 1994: chapter 9)
- Phoneme splits (Labov 1994: chapter 11)

4.1 Computational architecture

This computational system illustrates the emergence and shift of groupings of tokens at two nested levels of organization, corresponding abstractly to a possible lexical system with only one sublexical category level. For ease of exposition, I will refer to the larger level as *word*, and the lower level as *segment*. The architecture involves two agents in conversation, taking turns producing and categorizing words under the influence of biases arising in production and categorization biases (Figure 2).

Each agent begins a run of the program with a lexicon of four word categories[6], each seeded with a set of exemplars of previously encountered tokens of that word. Each word exemplar consists of an ordered series of segment exemplars. My interest is to illustrate lexicon-influenced sublexical category variation rather than the *de novo* emergence of sublexical chunking *per se*, so the available space for segment exemplars is pre-divided into two labeled dimensions, which we can think of as corresponding to phonetically-based dimensions such as voice onset time (VOT) or tongue height (Lindblom et al. 1984). Segment dimensions are represented as arbitrary scales from 0–100, where individual segment exemplars map to a single point on the scale. Each segment exemplar therefore references two kinds of information: a dimensional category label (e.g. tongue height), and a point on that dimension (e.g. a target tongue height position). Word exemplars correspondingly map to points in an n-dimensional space defined by their sequence of segment exemplars, e.g. [12, 20]. This multi-level exemplar structure (Wedel 2009; Walsh et al. 2010) allows the computational system to record and respond to the distribution of exemplar values of a particular segment both within word categories and also across all words. As a running metaphor to clarify the illustrations to follow, I will refer to the first dimension as representing an arbitrary VOT scale from [b] to [p], and the second dimension as representing an arbitrary tongue height scale from [a] to [i]. A two-segment word token with the values [12^{VOT}, $20^{TongueHeight}$] would therefore correspond in this metaphor to something in the neighborhood of [ba]. The difference between neighboring points is

---

6 The size of the lexicon can be significantly increased without a qualitative change in the system behaviors described below. I limit the lexicon to four words here to allow clear presentation of results in graphical form.
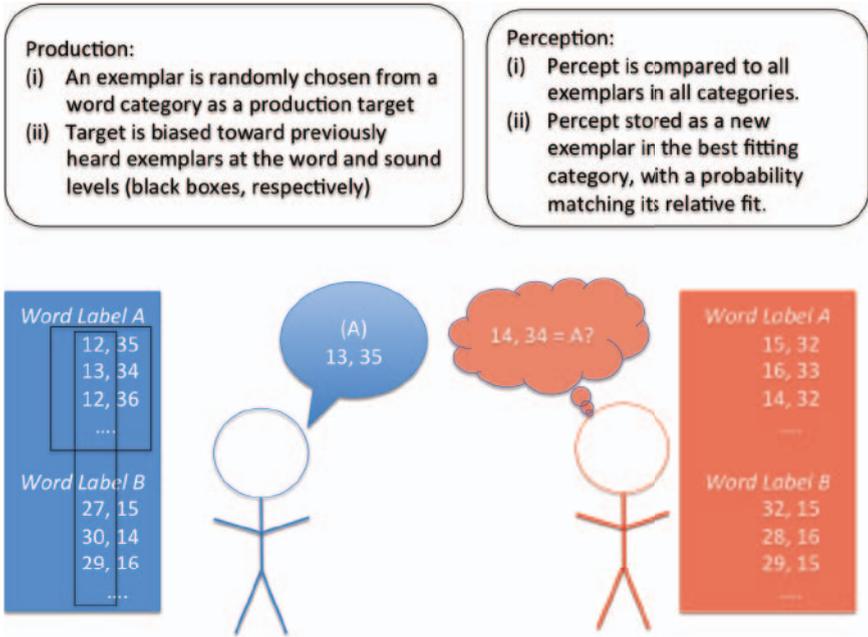
Fig. 2: The flow of information in the computational system.

small relative to the variation introduced by noise and the granularity of categorization, so these scales are functionally smooth (Pierrehumbert 2001).

For the agent whose turn it is to produce a given word, production begins with the random choice of a word exemplar from that category as an output target where the probability of choice is proportional to exemplar activation level (Hintzmann 1986; activation is calculated as an exponential function of recency, where exemplars that were stored 100 rounds previously have an activation level that is approximately .1% that of a new exemplar). To implement a within-word-category similarity bias, the segment exemplar values of this initial word target are stochastically biased toward the value at the same positions in all the word exemplars within the category. To implement a within-segment-dimension similarity bias, each individual segment exemplar value in the target is also stochastically biased toward all other segment exemplars that reference the same dimension across the entire lexicon[7]. For example, a target output value on the

---

7 To model similarity bias, a population vector is produced in relation to the target exemplar over all exemplars in that category. This vector is a weighted average of all exemplars mapped to the category, where both the Euclidean distance from the target exemplar and activation

VOT dimension is influenced by every single VOT exemplar token in memory, regardless of word category. Because influence is in relation to similarity, a VOT exemplar token in another word category that is similar to the output target will influence it more than one that is dissimilar.

To introduce variation, noise is also added to values of the output target by adding a normally distributed random value. This random value is biased slightly toward the center of the dimension, (i.e. a scale value of 50), in a simple model of production-based lenition (Pierrehumbert 2001; see also e.g. Lindblom et al. 1984 for arguments that the packing of phoneme inventories is in part a consequence of effort-minimization processes). The results described below do not depend on this lenition bias, but they contribute to the illustration by imposing a tendency for each segment exemplar distribution to drift toward the center of each dimension which encourages category merger (see discussion below).

The agent who is currently in the role of the listener then begins the categorization process by calculating the similarity of the speaker output to each category's stored word exemplars given their activations, in a variant of the Generalized Context Model (Nosofsky 1988). The overall similarities of the speaker output to each category are interpreted as a relative goodness of fit, and the speaker output is then stored as a new exemplar in the best fitting category. After all the speaker's word categories have been produced and categorized by the listener, roles reverse. This steady accumulation of new exemplars and decay of old ones allows for slow change in the system in response to noise and biases in production and categorization.

What do we expect then, given the three types of production bias that are built in to this system?

i.   Similarity bias at the word level
ii. Similarity bias at the segment level
iii. Lenition bias toward the center of each segment dimension

All else being equal, (i) the similarity bias at the word level will tend to keep tokens of a given word similar to one another; (ii) the similarity bias at the segment level will tend to pull segment tokens on a given dimension closer together across the lexicon, and (iii) the lenition bias will tend to pull segment tokens toward the center of their dimension. As a result, we expect that over sufficient rounds, all segment exemplars will come to cluster around a value of 50, resulting

---

(i.e. recency) influence each exemplar's contribution (cf. Nosofsky 1988). For more details on the computational architecture, see the Appendix.

in a lexicon of four homophones. In Section 2.3. above, we saw that the linguistic model without any contrast-maintenance effect predicted the same loss of global contrast.

To include a lexical contrast maintenance effect in the computational system, I include a categorization bias for greater word contrast through a simple version of a listener-based subtractive mechanism described in Section 2.3.1. When a speaker output is assigned to the best fitting category, it has a probability of being stored as a new exemplar that is the same as its calculated relative fit that category. For example, if the relative fit of a token to the best fitting category is .9, it has a 90% chance of being stored and a 10% chance of being discarded. Because a speaker output cannot influence future pronunciations if the listener does not store it, this bias against ambiguous speaker outputs allows more contrastive speaker outputs to have a greater influence on the trajectory of change in the evolving system. Note that this subtractive mechanism in categorization in the computational system is not to be taken as an argument that this is the specific pathway underlying a lexical contrast maintenance effect in *language*. It is used in this computational system because it is the conceptually least complex way to introduce a pressure for contrast maintenance at the higher 'word' level of organization. To help make this clear, I will refer to this specific mechanism in this computational system as a *bias against ambiguity*[8], rather than the more neutral term *lexical-contrast maintenance effect* that I use when referring to language.

4.2 Phoneme inventories and sublexical category contrast

To illustrate the effect of a competition between a segment-level bias toward greater similarity and a word-level bias against ambiguity, we start with a lexicon containing four, two-segment words. Both agents' lexicons are pre-seeded with

---

8 A simple variant of this anti-ambiguity bias is one in which every speaker output is stored as an exemplar in the best fitting category, but where its *activation* value is proportional to its relative similarity score to that category (see Barr 2004 for a similar implementation). More ambiguous speaker outputs have a lower relative similarity score to their best fitting category and are stored with a lower initial activation level, and therefore contribute less to the future behavior of the system. Speaker-based contrast maintenance algorithms are also possible which model a speaker-based hypothesis for hyperarticulation (see Section 2.3.1. above), for example by exaggerating speaker outputs in relation to similarity to stored exemplars in competing word categories in the speaker's lexicon. These are computationally more complex, so for simplicity of exposition, I have chosen an analogue of the listener-based hypothesis for hyperarticulation for this computational system.
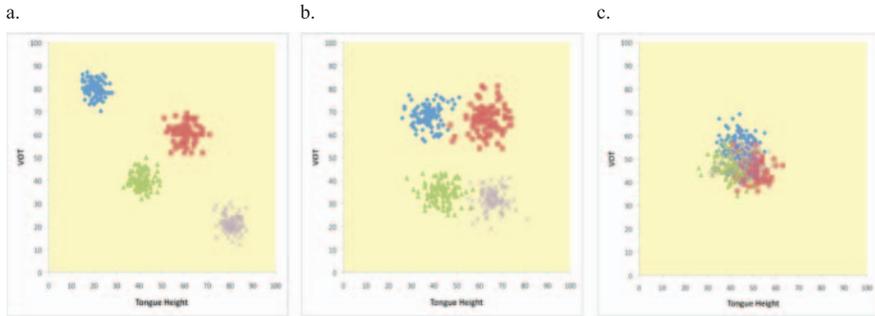
a.   b.   c.

Fig. 3: Evolution of a 4-word lexicon in the absence of any anti-ambiguity bias.

100 clustered exemplars of each word category. I intentially use a simple word-structure where each exemplar is composed of one point on two separate dimensions, because this allows us to represent each exemplar in an easily visualizable 2-dimensional graph in which the y-axis represents the first dimesion and the x-axis the second. In keeping with our descriptive metaphor, the y-axis is labeled 'VOT' and the x-axis 'Tongue Height' (Figure 3). Each point on the graph therefore represents a single word exemplar with corresponding VOT and Tongue Height values. The important thing to note about the initial lexicon shown in Figure 3a is that it doesn't have what a we would recognize as a phoneme inventory. Each word category contains segment exemplars whose value-ranges are different from every other word on that dimension, with the result that there are as many VOT and Tongue Height value-clusters as there are words. Because there is no coherent system of sublexical units shared across words, this initial lexicon does not exhibit duality of patterning.

In a first control run of the experiment, we eliminate any word-based bias against ambiguity by requiring the listener to store every speaker output in the best-fitting category no matter how ambiguous. Figures 3a to 3c show the progression of one agent's lexicon over the course of 4000 rounds under these conditions[9]. Note that by 1500 rounds (Figure 3b), similarity bias at the segment level drives the development of a compact inventory of segment exemplar distributions: each distribution of sound exemplars is shared between words, as can be seen in the roughly square arrangement of the word exemplar clouds in the two-dimensional space. However, in Figure 3c we see that the same similarity bias

---

9 The two agents' lexicons remain tightly linked over all simulation conditions, so only one is shown.

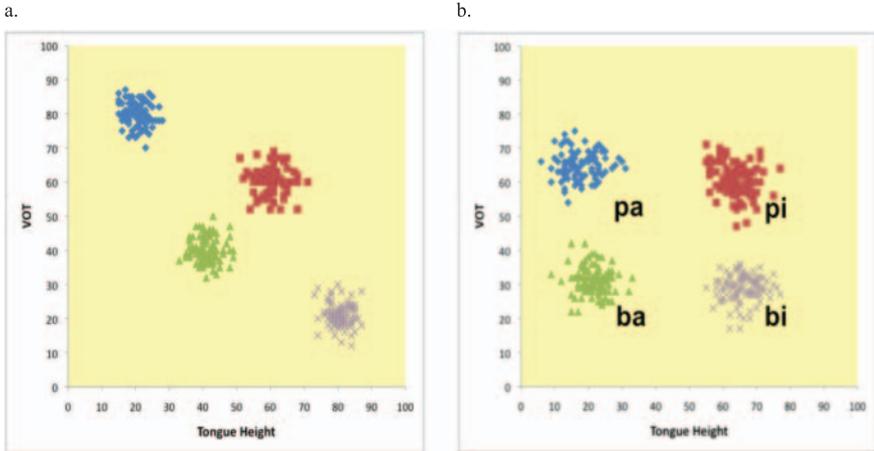a.   b.

Fig. 4: Evolution of a four-word lexicon with a word-level anti-ambiguity bias.

eventually drives overlap of all segment exemplar distributions, resulting in the collapse of word distinctions – the lexicon is now as compact as it can get, given the parameters of the system. Compare this transition to the transition above in Figures 1b to 1c: the process of progressive, similarity bias-driven collapse of all distinctions is the same in both systems.

Figure 4 shows the starting and end state of a representative run of the same lexicon after 4000 rounds in which the bias against ambiguity at the word level is restored. In this case, we see that the distributions of word-exemplars shift so that they line up vertically and horizontally in the graph, corresponding to having developed a more compact system in which segments are shared across words. However, full collapse of segment distinction is inhibited by the anti-ambiguity effect in word-exemplar storage. In this system, similarity bias at the segment level and anti-ambiguity bias at the word level work against one another, and the solution to this cross-level conflict is a lexicon of distinct words built from a compact inventory of segments shared between words. This allows us to describe these words with a more compact symbolic system, e.g. /bi, pi, ba, pa/.

The results shown here are representative. To demonstrate this, I ran twenty independent runs of the simulation starting with fully randomized word-exemplars in each category (see the second syllable in Figure 6a below for an example of a random starting state). Ten of these were run without the anti-ambiguity bias, and ten with the anti-ambiguity bias. For the ten without the bias, each lexicon collapsed (as measured by majority overlap of all categories) by an average of 3100 rounds, with a maximum of 5000 rounds. In ten runs of 10,000

rounds including the anti-ambiguity bias, there were no instances of collapse[10]. Seven of the ten runs with the anti-ambiguity bias resulted in square configurations as in Figure 4b, and three resulted in other vertically and horizontally arranged configurations such as a 'T' or 'L'. What all these configurations have in common is the sharing of sublexical exemplar distributions between lexical categories.

It is worth emphasizing that there is nothing in the computational system architecture that directly acts to inhibit segment category merger. Instead, segment distinctions are maintained indirectly through a mechanism that inhibits *word* category merger. This is parallel to the biological example discussed above, in which selection at the level of the individual passes on genes to a future generation in an all-or-nothing fashion, yet over time results in changes in particular gene frequencies in the population. Likewise, the contrast-maintenance mechanism in this example system only directly influences the storage probability of word exemplars, not segment exemplars. When multiplied over many exchanges, however, this word-based mechanism can influence change in individual segment categories. For example, when a pair of words is distinguished primarily by the values of a particular segment pair (e.g. when that segment pair functionally defines a minimal word pair), less contrastive tokens of either of these segments are likely to result in less contrastive tokens of those words. Our computational word-contrast maintenance mechanism operates by storing less contrastive word tokens less often, with the result that averaged over many rounds, segment variants that are statistically associated with lower word contrastiveness will contribute correspondingly less to the distribution of segment exemplar values. Given the subtractive contrast maintenance algorithm implemented here, this pathway relies on the steady generation of a range of segment variants through random variation. If a particular segment variant contributes more to the contrast of the output word token it finds itself in, it is more likely to be stored by the listener, and thereby incrementally shift the distribution of variation in the set of exemplars in memory toward greater contrast.

---

10 If we set the time of collapse for each of the runs with the anti-ambiguity bias as the last round, i.e. 10000, the means of collapse times for the two conditions are significantly different as assessed using the unequal variance t-test ($t(9) = 21.7$, $p < .05$). The small lenition-bias toward the center of each dimension was included in part to speed collapse by encouraging sound-exemplar distributions to drift close enough for similarity bias to promote further collapse. If this lenition bias is not included, the mean and variance of time to collapse is greater, but collapse remains inevitable. Effort-minimization-based lenition biases are theoretically and experimentally supported, but here the lenition bias plays an additional heuristic role in the illustration by making the *lack* of collapse in the anti-ambiguity condition more informative.
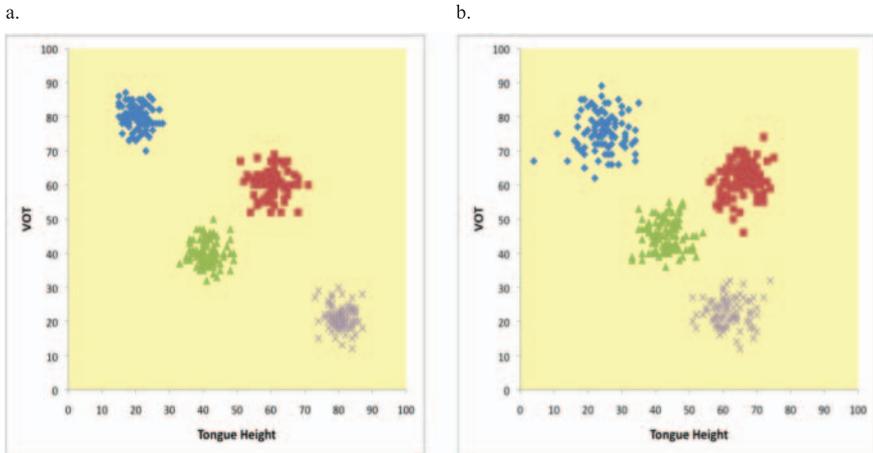
a.

b.

Fig. 5: Without any similarity bias along segment dimensions, words remain contrastive at 4000 rounds but do not share a segment inventory.

Finally, the development of a system of segment categories shared between words requires a pathway for segments in different words to become more similar. To illustrate this, Figure 5 shows a run analogous to that in Figure 4, but in which similarity bias at the segment level is turned off. This eliminates the between-word influence of segment-exemplar distributions on one another. In the results of this run, we can see that word-level bias against ambiguity keeps word categories distinctive, preventing the collapse seen in Figure 3c. However, because there is no pathway for segment variants to spread across words, each word category retains an idiosyncratic distribution of segment exemplars.

The simple illustrations above showed that in this system, initially distinct distributions of segment exemplars collapse when that distinction does not contribute to word contrast. But as discussed above with the examples of *pat* vs *passion*, individual sublexical category tokens may not contribute substantially to contrast in the word they find themselves in and yet retain their categorial identity. Speech production models that include mutual influence of detail stored at varying levels of organization are specifically able to account for lexically specific variation in segment pronunciations (reviewed in Walsh et al. 2010). Consequently, this type of model could in principle accommodate a variant of English in which every initial labial stop that is not required for lexical contrast to be pronounced with a VOT intermediate between that for [p] and [b]. Instead, complete neutralizations of category contrast over the course of language change tend to be sublexically coherent, that is, to be associated with a delimited sub-

lexical context and to hold in all lexical items that contain that context. For example, the /ɪ~ɛ/ vowel contrast has merged in dialects of Southern American English, but only in the context of a following coda nasal (i.e. the 'pin~pen' merger (Labov et al. 2006); see Blevins and Wedel 2009 for a discussion of paradigmatic anti-homophony patterns).

The Network Feedback Model described here (and argued for in more detail in Wedel 2007) accounts for the strong tendency for sublexical category patterns to be regular across the lexicon through similarity bias in the production and perception of sublexical category tokens, where categories of multiple and overlapping sizes can exist simultaneously[11]. We can use our computational system to show how a similarity bias and a lexical contrast maintenance effect can work together to create a constrained system of sublexical categories, despite the fact that each category contributes to lexical contrast in only a subset of the lexemes it appears in. We start with four word categories as before, but now each has four segments rather than two, arranged as two sequential 'syllables' of VOT-Tongue Height pairs. We start the run with the initial syllables arranged in the configuration shown in Figure 4a, which we know tends to evolve to the square, maximally compact configuration seen in Figure 4b. Because the four words are already fully distinguished from each other by their initial syllables, the segment values in the second syllables of these words could be anything at all without impacting word token contrast for the listener in this computational system. To enable us to better see how the segment values in the second syllables evolve, the VOT and Tongue Height segments in the second syllable of each exemplar are seeded with random values between 0–100 at the beginning of the run. In this way, the second syllables sample the entirety of each dimension at the start. Figure 6a shows this starting state, and Figure 6b the state of the lexicon after 4000 rounds, where the first graph in each case represents the initial syllable of each word, and the second the second syllable. Because the second syllable is redundant for word-contrast, we might expect these segments to collapse to the middle, as in Figure 3c. Instead, the segment distributions in the second syllable have collected in the attractors formed by the already established segment distributions in the first syllable. This system

---

11 Note that this computational system does not explicitly represent sublexical category labels (e.g. /b/ versus /p/). This is not a claim that lexical production and perception cannot proceed in whole or in part through the processing of explicit, learned sublexical category labels. Rather, this model proposes that produced and perceived variation is biased toward local maxima at the sublexical level. This encourages the development of coherent sublexical patterns across the lexicon, and thereby may feed into the abstraction of sublexical categories from input data in acquisition (Pierrehumbert 2003).
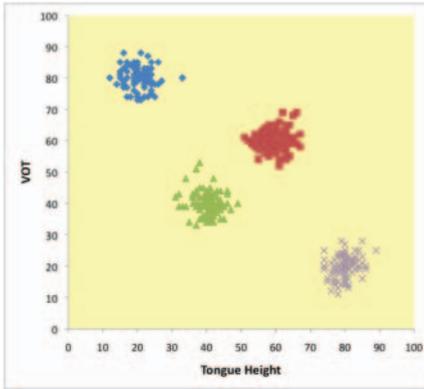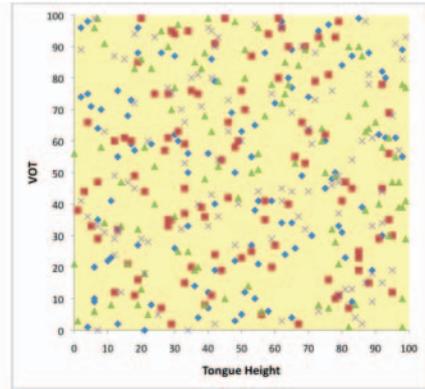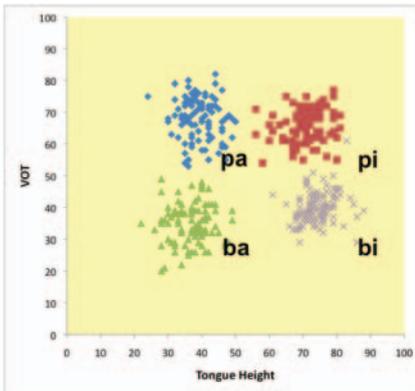
syllable 1　　　　　　　　　　　　　　　　　　syllable 2

Fig. 6a: Similarity bias at the segment level promotes system coherence. a. At the start, the first syllable of each word carries all necessary contrast, and the second syllable consists of random segment exemplars.
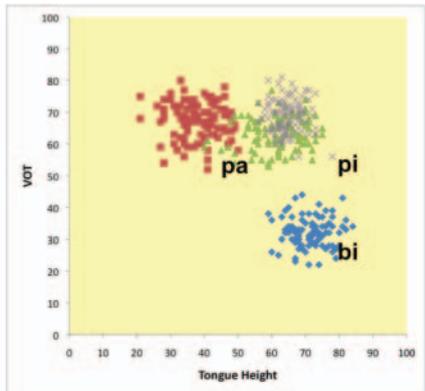
syllable 1　　　　　　　　　　　　　　　　　　syllable 2

Fig. 6b: The first and second syllables after 4000 rounds.

remains stable due to a strong segment-level similarity bias, despite the fact that the second-syllable segments contribute little to word contrast in this computational system. Within the Network Feedback Model, the ability of previously established peaks in a distribution to act as attractors in change can be understood as a simple model of the tendency to structure-preservation in language change (discussed in Blevins 2009).

4.3 Chain shifts and lexical diffusion

Chain-shifts are parallel changes in the phonetic realizations of adjacent phoneme categories that maintain phonetic distance between those categories (Labov 1994: chapter 9). As detailed in Section 2.2. above, we expect variants to spread and consolidate across categories given a similarity bias on variation operating within a rich-memory, multi-level categorial system. The model described here provides a pathway for chain shifts to occur indirectly through the influence of sublexical variants between lexical categories (cf. lexical diffusion (Wang 1969; Phillips 2006)). To provide a simple illustration, we start again with a four-word CV lexicon with the word tokens arranged in the square that represents the most stable state in the system (see Figure 4c above). I will refer to these words by their approximate starting pronunciations in our metaphor: *ba*, *bi*, *pa*, *pi*. In this run, I introduce an additional bias on the y-axis (VOT) dimension toward the scale center at 50, but only for output tokens of the word *ba*, seen in the bottom left corner of the graph. Figure 7 shows the state of the system after 4000 rounds, with arrows to show the relative movement of each category. When we look at the time course of category center movement on the VOT scale, we see that over the course of about 1000 rounds, the distribution of values on the VOT dimension for tokens of the word *ba* slowly creeps from its original center around 35 up to 50, as we would expect given the externally imposed bias. Interestingly, the other three words shift in concert as well with just a short lag. Given that the original VOT distance between *ba* and *pa* is optimal given the competing demands of the lenition bias and the anti-ambiguity bias on word storage, we expect that *pa* should shift upwards as *ba* encroaches. At the same time, we expect that the VOT values for *bi* might rise as well, because VOT token values for *bi* and *ba* are already very similar, with the result that they all influence one another strongly in production. Finally, the distribution of *pi* tokens shifts upwards as well, both through the similarity bias from the shifting distribution of VOT token values in *pa*, and through the anti-ambiguity bias which drives it away from *bi*. Note that the influence between *ba* and the other words is not one-way: similarity and anti-ambiguity biases provide a tight constraining connection between all categories in the system, such that even as the introduced VOT bias is continually creating higher *ba* tokens, the similarity and anti-ambiguity biases arising from the rest of the lexicon push them back. This inertia can be seen from the fact that when *ba* is the only word in the lexicon, the shift from a mean position of 35 on the VOT dimension to 50 is completed in just a few hundred rounds, rather than in more than 1000 (not shown).

    This example illustrates a change in the system that is analogous to a push chain-shift, in which a shift of one sublexical category towards another along a
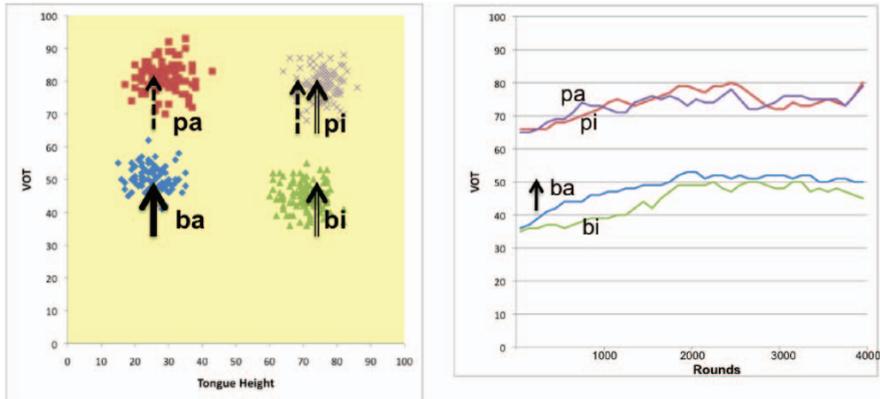
Fig. 7: The spread of a chain-shift from word pairs to the larger lexicon. The run begins with exemplars in four-word categories centered at [35, 35], [35, 65], [65, 35] and [65, 65] (represented by the starting points of the arrows). After 100 rounds, an external production bias (black arrow) is introduced on the initial (VOT) values of the word category centered at [35, 35], labeled *ba*, which shifts the average values slowly toward 50. This gradual shift in the average value of VOT exemplar values in this category is shown in the right-side of the graph over the course of 4000 rounds. The average value of the VOT exemplars in the other word categories shifts upwards as well. The heavy arrow denotes the the externally imposed bias, which has two distinct subsidiary effects within the system. The dashed arrows represent pressure in the system for change through anti-ambiguity bias from the encroachment of one word toward another, and the outlined arrows represent pressure in the system for change through similarity bias at the segment level, i.e. lexical diffusion, which acts to consolidate segment exemplar distributions across words.

shared dimension results in a compensating shift of the other. In chain-shifts, contrast can be maintained by shifting further away on the same dimension and/ or by shifting the burden of contrast to a different dimension (see the Pittsburgh English example below; Labov 1994: chapter 9; Maclagan and Hay 2007; see also Ettlinger 2007). In this computational system, we see how a pre-existing, stable sound system can shift as a whole through pressure only in one part. More broadly in the Network Feedback Model, stable categorial relationships tend to remain in place through the interaction of (i) pattern spread and maintenance through similarity bias and (ii) lexical contrast maintenance effects. As a consequence, the model predicts that any process (e.g. sociological factors, Labov 2001) that induces a particular shift in the pronunciation of some words has the ability to indirectly promote a corresponding change across the lexicon, and with it a larger chain of shifts in the sublexical category contrast system.

4.4 Cue shifts and phoneme splits

In the previous section, we used this computational system to illustrate the way competition between a similarity bias and lexical contrast maintenance can shift the range of sublexical variants along a single dimension of sublexical contrast. In this section, I review arguments previously made in Wedel (2006: 261–269) that the same processes can also promote shifts in cue robustness *between* dimensions of contrast. Sublexical category identity is often predicted by multiple phonetic cues (e.g. Whalen 1981; Whalen et al. 1993; McMurray et al. 2002; Beddor 2009), and that listeners can shift their attention between cues depending on the degree to which a cue is perceptible (Repp 1982; Beddor 2009). Cues to a sublexical category can be simultaneous (or otherwise hosted by the same segment), for example in the formant and vowel length cues that both contribute to distinguishing tense/lax English vowel pairs like /i/ and /ɪ/. However, cues to the identity of a particular sublexical category can also be found elsewhere in the utterance, often on a neighboring segment. For example, the voicing distinction in coda obstruents in English is secondarily cued by preceding vowel length, in addition to the primary cue of voicing of the obstruent itself (Raphael 1971). Through the course of language change, the burden of identifying a category can shift between cues, such that a primary cue to a sublexical category can be lost while an originally secondary cue becomes more robust. The elevation of an originally secondary cue into a new phonemic contrast is termed a split (reviewed in Labov 1994: chapter 11). For example, Labov and colleagues report that in Pittsburgh English tokens of the vowel categories /ʌ/ and /aw/ have shifted together such that their formant values now largely overlap. However, the originally nonphonemic length difference between the shorter /ʌ/ and longer /aw/ has become categorical at least for some speakers of this dialect (Labov et al. 2006; see also Labov and Baranowski 2006). Examples of splits arising through a shift in cue strengths across segments include the emergence of contrastive vowel nasalization in French upon loss of coda nasals (reviewed in Beddor 2009), and the proposed development of contrastive tone on vowels through the loss of stop voicing distinctions in onsets (Hombert et al. 1979).

From the point of view of the lexicon, chain shifts and splits produce the same result: in both, phonetic change occurs in a way that preserves lexical distinctions. I showed above how a lexical contrast maintenance effect can account for chain shifts in the context of a similarity bias promoting pattern coherence. Similarly, because contrast maintenance in the model operates to support contrast between *lexical* categories rather than *sublexical* categories, the model predicts that a contrast maintenance effect can exploit pre-existing correlations between two cues to promote the enhancement of one cue in compensation for

the loss of the other. This occurs through shifts in the distribution of cue variants represented in lexical exemplars, just as we saw in the section above in the illustration of chain-shifts. As an example, consider the relationship between coda obstruent voicing and preceding vowel length in English: relatively greater vowel length is associated with voicing in a following obstruent (compare [pæt] vs [pæːd]), and English speakers use this cue in lexical identification (Raphael 1971). The Network Feeback Model predicts that if the glottal pulsing cue to voicing of coda obstruents weakens and if there are many lexical competitors that depend on this cue, lexical contrast maintenance should promote enhancement of any other existing cues to lexical identity. Consistent with this, a study of American English spoken in Watertown, Wisconsin (Purnell et al. 2005) shows that the glottal pulsing cue distinguishing voiced/voiceless coda obstruents has weakened, presumably under the influence of German, and that vowel length has in turn become the primary cue to lexical identity for the words previously distinguished by the coda obstruent voicing distinction (for more discussion of this model's predictions concerning phonemic splits and an associated computational model, see Wedel 2006: 261–269).

4.5 Further issues

A number of parameters of the overall problem of contrast maintenance in this model remain unfixed. For example, what levels of description are most relevant for understanding the effect of ambiguity in usage on sound change? Wedel et al. (In press) report that for their dataset, ambiguity at the root level accounts for more of the data than ambiguity at the phoneme or the surface word form levels, but more work needs to be done to establish this. Further, at what level(s) does frequency play a role? Greater frequency of use of a category is predicted to allow faster change in response to a bias (Bybee 2001, 2002a; Hay and Maclagan 2012; Wedel et al. In press). However, greater frequency of use also correlates with lower uncertainty (Shannon 1948). As a result, high frequency words should experience relatively less competition from minimal pair neighbors – at the same time that their greater frequency allows them to change more rapidly in response to that competition. As a result, the model predicts that frequency may be correlated with opposing tendencies depending on the context in which it is assessed.

What about language usage versus acquisition? The model, as it stands, accounts for a variety of types of sublexical category change through contrast maintenance between established lexical categories. As such, it predicts that measures

of language usage among post-acquisition speakers will better account for these types of changes than measures in language acquisition. Finally, the model does not make specific predictions concerning the influence of network structure on sublexical category change in a larger speech community. All of these remain exciting areas for further exploration.

5 Discussion

In traditional phonological theory, sublexical categories are atomic primitives of the language system, and lexemes are composed from these categories because they *must* be. In the model described here, lexical categories tend to be characterized by shared sublexical categories because they *can* be. Hockett (1960) argued that in any communication system, including human language, there is no intrinsic reason why form-meaning categories such as words must be composed from an inventory of non-meaningful parts. Instead he proposes that in human communication systems with large meaning spaces, compositional structure tends to emerge as a response to the cognitive burden of perceiving and producing a large number of holistic and gradiently distinctive forms. More recently, a growing body of computational and experimental work has added nuance to this proposal by showing that initially holistic systems spontaneously develop compositional structure when subjected to cycles of imperfect production, categorization and learning (Kirby 1999; de Boer 2001; Brighton et al. 2005; Wedel 2007; Griffiths and Kalish 2007; Cornish et al. 2009; Verhoef and de Boer 2011, and many more). This research program explores the larger hypothesis that much of the structure that we see in language, and in its change over time, arises through interacting pathways of information transformation within and across individuals and generations (e.g. Christiansen and Chater 2008).

Sandler et al. (2011) present intriguing evidence consistent with the notion that form-meaning categories are causally prior to sublexical categories in an ongoing study of the properties and development of Al-Sayyid Bedouin Sign Language (ABSL). This language has emerged historically recently within an isolated community, and Sandler and colleagues show that most signers do not show evidence of sublexical structure in their signs. Nonetheless, Sandler and colleagues present evidence that ABSL functions as a full language, supporting Hockett's claim that duality of patterning is not a prior necessity for language to emerge. Additionally however, they report that signers from a family with many deaf members do show evidence for some categorical patterns of handshape assimilation, as well as instances where ease of production alters formational elements

resulting in a less iconic image. The fact that these alternations are not shared across the community indicates that they are not simply a consequence of motor-bias, and the observation that the assimilated handshape does not convey independent meaning suggests that it is now a conventionalized subpart of a larger sign. On the basis of these observations, Sandler et al. (2011) suggest that although lexical systems can function in the absence of sublexical structure, general cognitive, and motoric biases may promote the development of a corresponding sublexical category system through progressive conventionalization of sub-parts of originally holistic forms.

Since Baudouin de Courtenay (1895), a recurring *leitmotif* in phonological research has been that biases in production, perception and learning operate continually to influence the range of variants that arise and propagate through a speech community, and that the synchronic properties of a given phonological system cannot be understood independently from these processes (e.g. Ohala 1989; Labov 1994, 2001; Bybee 2001; Blevins 2004, and many others). A corresponding current in my own work has been understanding how the interaction between particular types of biases can induce the self-organization of higher-order patterns within a lexicon over time. In particular, evidence for the storage and reproduction of experience variation predicts that given a highly networked lexicon structure, biases on the production and perception of individual words in usage can influence the long-term trajectory of change in the broader sublexical system (Bybee 2002a; Wedel 2007; Hay and Maclagan 2012).

In this and previous work (Wedel 2004a, 2007), I have argued that a general cognitive bias toward variants that are similar to previous experience can account for the emergence of conventionalized phonological behavior, given storage of phonetic detail in memory and a feedback loop between perception and production. A central prediction of this feedback-based model is that similarity is self-reinforcing: the more items that exhibit a pattern, the greater the resulting bias in variation toward that pattern. This self-referential property of the Nework Feedback Model provides general accounts for a range of phonological phenomena (Wedel 2007), but it also suggests that in the absence of countervailing support for contrast, all distinctions in a lexicon should inevitably collapse. However, linguistic, experimental and corpus evidence suggest that there do exist some mechanism(s) that promote contrast between lexical categories which compete in usage. My intent here has been to show that with the addition of this type of bias, the model provides a single general pathway for lexical contrast maintenance processes in communication to initiate and guide a range of phonological phenomena such as phoneme category mergers, splits, and chain-shifts. Under this account, the probability of these different types of sublexical category change should be influenced not only by the properties of the cues to a category relative

to the rest of the sublexical system, but also by the degree to which those cues distinguish lexical items in usage.

References

Aylett, M. & A. Turk. 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*, 81–104. Amsterdam: Elsevier.

Baese, M. & M. Goldrick. 2009. Mechanisms of interaction in speech production. *Language and cognitive processes* 24. 527–554.

Barr, D. J. 2004. Establishing a conventional communication system: Is common knowledge necessary? *Cognitive Science* 28. 937–62.

Baudoin de Courtenay, J. 1895. *Versuch einer Theorie phonetischer Alternationen: ein Capitel aus der Psychophonetik*. Strassburg, Germany/Crakow, Poland: Trübner. [English translation ('An attempt at a theory of phonetic alternations'). Published in E. Stankiewicz (ed.) 1972, *A Baudouin de Courtenay anthology: The beginnings of structural linguistics*, 144–212. Bloomington, IN: Indiana University Press.

Beckner, C., R. Blythe, J. Bybee, M. H. Christiansen, W. Croft, N. C. Ellis, J. Holland, J. Ke, D. Larsen-Freeman & T. Schoenemann. 2009. Language is a complex adaptive system. *Language Learning* 59. 4S1.

Beddor, P. S. 2009. A coarticulatory path to sound change. *Language* 85(4). 785–821.

Blevins, J. 2004. *Evolutionary phonology*. Cambridge: Cambridge University Press.

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*, 33–49. Canberra: Pacific Linguistics.

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

Brighton, H., S. Kirby & K. Smith. 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,* 291–309. Oxford: Oxford University Press.

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

Bybee, J. 2002a. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14. 261–290.

Bybee, J. 2002b. Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition* 24. 215–221.

Bybee, J. & J. L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22. 381–410.

Christiansen, M. H. & N. Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31(5). 489–509.

Cornish, H., M. Tamariz & S. Kirby. 2009. Complex adaptive systems and the origins of adaptive structure: What experiments can tell us. *Language Learning* 59(s1). 187–205.

Currie-Hall, K. 2009. *A probabilistic model of phonological relationships from contrast to allophony.* The Ohio State University doctoral dissertation.

de Boer, B. 2001. *The origins of vowel systems*. Oxford: Oxford University Press.

de Boer, B. 2006. Computer modelling as a tool for understanding language evolution. In N. Gontier, J. P. Van Bendegem & D. Aerts (eds.), *Evolutionary epistemology, language and culture,* 381–406. Dordrecht: Springer.

Dell, G. S. 1986. A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93. 283–321.

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,* 195–223. Cambridge, MA: MIT Press.

Ernestus, M. 2011. Gradience and categoricality in phonological theory. In M. Van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell companion to phonology*, 2115–2136. Oxford: Blackwell.

Ettlinger, M. 2007. Shifting categories: An exemplar-based computational model of chain shifts. *Proceedings of the 29th Annual Conference of the Cognitive Science Society*, 239–244. Austin, TX: Cognitive Science Society.

Frisch, S. 1996. *Similarity and frequency in phonology*. Chicago: Northwestern University doctoral dissertation.

Ganong, W. F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6. 110–125.

Gessner, S. & G. Hansson. 2004. Anti-homophony effects in Dakelh (Carrier) valence morphology. In M. Ettlinger, N. Fleischer & M. Park-Doob (eds.), *BLS 30: Proceedings of Berkeley Linguistic Society,* 93–104. Berkeley: Berkeley Linguistics Society.

Gilliéron, J. 1918. *Généalogie des mots qui désignent l'abeille d'après l'atlas linguistique de la France*. Paris: E. Champion.

Goldinger, S. D. 2000. The role of perceptual episodes in lexical processing. In A. Cutler, J. M. McQueen & R. Zondervan (eds.), *Proceedings of SWAP (Spoken Word Access Processes)*, 155–158. Nijmegen: Max-Planck-Institute for Psycholinguistics.

Gordon, M. J. 2002. Investigating chain shifts and mergers. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (eds.), *Handbook of language variation and change*, 244–266. Oxford: Blackwell.

Griffiths, T. L. & M. L. Kalish. 2007. Language evolution by iterated learning with Bayesian agents. *Cognitive Science* 31. 441–480.

Guenther, F. H. & M. N. Gjaja. 1996. The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100. 1111–1121.

Hare, M. & J. L. Elman. 1995. Learning and morphological change. *Cognition* 56. 61–98.

Harrington, J., S. Palethorpe & C. I. Watson. 2000. Does the Queen speak the Queen's English? *Nature* 408. 927–928.

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

Hintzman, D. 1986. "Schema abstraction" in a multiple-trace memory model. *Psychological Review* 93. 411–428.

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

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

Hombert, J.-M., J. J. Ohala & W. G. Ewan. 1979. Phonetic explanations for the development of tones. *Language* 55. 37–58.

Hume, E. & K. Johnson. 2001. A model of the interplay of speech perception and phonology. In Hume & Johnson (eds.), *The role of speech perception in phonology*, 3–26. New York: Academic Press.

Joanisse, M. F. & M. S. Seidenberg. 1997. [i e a u] and sometimes [o]: Perceptual and computational constraints on vowel inventories. In M. Shafto & P. Langley (eds.), *Proceedings of the 19th Annual Meeting of the Cognitive Science Society*, 331–336. Hillsdale, NJ: Erlbaum.

Johnson, K. 1997. Speech perception without speaker normalization. In K. Johnson & J. W. Mullennix (eds.), *Talker variability in speech processing*, 145–166. San Diego: Academic Press.

Ju, M. & P. A. Luce. 2006. Representational specificity of within-category phonetic variation in the long-term mental lexicon. *Journal of Experimental Psychology: Human Perception and Performance* 32. 120–138.

Kaplan, A. 2011. How much homophony is normal? *Journal of Linguistics* 47(3). 631–671.

Kenstowicz, M. 1994. *Phonology in generative grammar*. Oxford: Blackwell.

King, R. 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.

Kraljic, T. & A. Samuel. 2005a. Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review* 13. 262–268.

Kraljic, T. & A. Samuel. 2005b. Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology* 51(2). 141–178.

Kuhl, P. K. 1991. Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50. 93–107.

Labov, W. 1994. *Principles of linguistic change, Volume 1, Internal Factors*. Oxford: Blackwell.

Labov, W. 2001. *Principles of linguistic change, Volume 2, External Factors*. Oxford: Blackwell.

Labov, W., S. Ash & C. Boberg. 2006. *Atlas of North American English: Phonetics, phonology, and sound change*. Berlin: Mouton de Gruyter.

Labov, W. & M. Baranowski. 2006. 50 milliseconds. *Language Variation and Change* 18. 1–18.

Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H and H theory. In W. Hardcastle & A. Marchal (eds.), *Speech production and speech modeling*, 403–439. Dordrecht: Kluwer.

Lindblom, B., P. MacNeilage & M. Studdert-Kennedy. 1984. Self-organizing processes and the explanation of language universals. In B. Butterworth, B. Comrie & Ö. Dahl (eds.), *Explanations for language universals*, 181–203. Berlin: Mouton de Gruyter.

Luce, P. A. & D. B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1–36.

Maclagan, M. & J. Hay. 2007. Getting *fed* up with our *feet*: Contrast maintenance and the New Zealand English 'short' front vowel shift. *Language Variation and Change* 19(1). 1–25.

Maddieson, I. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

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

Martinet, A. 1960. *Éléments de linguistique générale*. Paris: Armand Colin.

McMillan, C. T. & M. Corley. 2010. Cascading influences on the production of speech: Evidence from articulation. *Cognition* 117. 243–260.

McMurray, B., M. Tanenhaus & R. Aslin. 2002. Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86. B33–B42.

Mielke, J. 2008. *The emergence of distinctive features*. Oxford: Oxford University Press.

Munson, B. 2007. Lexical access, lexical representation, and vowel production. In J. S. Cole & J. I. Hualde (eds.), *Laboratory phonology* IX, 201–228. Berlin: Mouton de Gruyter.

Munson, B. & N. P. Solomon. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47. 1048–1058.

Nielsen, K. 2007. The interaction between spontaneous imitation and linguistic knowledge. *UCLA Working Papers in Phonetics* 105. 125–137.

Norris, D., J. M. McQueen & A. Cutler. 2003. Perceptual learning in speech. *Cognitive Psychology* 47. 204–238.

Nosofsky, R. 1988. Similarity, frequency, and category representations. *Journal of Experimental Psychology, Learning, Memory and Cognition* 14. 54–65.

Ohala, J. 1989. Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (eds.), *Language change: Contributions to the study of its causes*. [Series: *Trends in Linguistics, Studies and Monographs* No. 43], 173–198. Berlin: Mouton de Gruyter.

Oudeyer, P.-I. 2002. Phonemic coding might be a result of sensory-motor coupling dynamics. In B. Hallam, D. Floreano, J. Hallam, G. Hayes & J.-A. Meyer (eds.), *Proceedings of the 7th international conference on the simulation of adaptive behavior,* 406–416. Cambridge, MA: MIT Press.

Pardo, J. S. 2006. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America* 119. 2382–2393.

Phillips, B. S. 2006. *Word frequency and lexical diffusion*. Bastingstoke: Palgrave Macmillan.

Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (eds.), *Frequency effects and the emergence of linguistic structure*, 137–157. Amsterdam: John Benjamins.

Pierrehumbert, J. 2002. Word-specific phonetics. In C. Gussenhoven & N. Warner (eds.), *Laboratory phonology* 7, 101–139. Berlin: Mouton de Gruyter.

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

Pierrehumbert, J. 2006. The new toolkit. *Journal of Phonetics* 34. 516–530.

Pisoni, D. B. & S. V. Levi. 2007. Some observations on representations and representational specificity in speech perception and spoken word recognition. In G. Gaskell (ed.), *The Oxford handbook of psycholinguistics*, 3–18. Oxford: Oxford University Press.

Purnell, T., J. Salmons & D. Tepeli. 2005. German substrate effects in Wisconsin English: Evidence for final fortition. *American Speech* 80. 135–164.

Raphael, L. 1971. Preceding vowel duration as a cue to the voicing characteristic of word-final consonants in American English. *Journal of the American Acoustical Society* 51. 1296–1303.

Repp, B. H. 1982. Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92. 81–110.

Sanchez, K. 2011. *Do you hear what i see? The voice and face of a talker similarly influence the speech of multiple listeners*. University of California at Riverside Doctoral dissertation.

Sancier, M. L. & C. A. Fowler. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25. 421–436.

Sandler, W., M. Aronoff, I. Meir & C. Padden. 2011. The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory* 29. 503–543.

Scarborough, R. A. 2004. Coarticulation and the structure of the lexicon. University of California Los Angeles Doctoral dissertation.

Scarborough, R. A. 2010. Lexical and contextual predictability: Confluent effects on the production of vowels. *Papers in Laboratory Phonology* 10. 557–586.

Shannon, C. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27. 379–423, 623–656.

Shattuck-Hufnagel, S. & D. H. Klatt. 1979. The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18. 41–55.

Silverman, D. 2010. Neutralization and anti-homophony in Korean. *Journal of Linguistics* 46(2). 453–482.

Smith, K., S. Kirby & H. Brighton. 2003. Iterative learning: A framework for the emergence of language. *Artificial Life* 9. 371–386.

Studdert-Kennedy, M. & L. Goldstein. 2003. Launching language: The gestural origin of discrete infinity. In M. H. Christiansen & S. Kirby (eds.), *Language evolution: The state of the art*, 235–254. Oxford: Oxford University Press.

Surendran, D. & P. Niyogi. 2006. Quantifying the functional load of phonemic oppositions, distinctive features, and suprasegmentals. In O. Nedergaard Thomsen (ed.), *Competing models of language change: Evolution and beyond*, 43–58. Amsterdam: John Benjamins.

Trubetzkoy, N. 1939. *Grundzüge der Phonologie*. Prague: Travaux du cercle linguistique de Prague.

van Son, R. J. J. H. & L. C. W. Pols. 2003. *How efficient is speech?* Paper presented at the Proceedings of the Institute of Phonetic Sciences, Amsterdam.

Verhoef, T. & B. de Boer. 2011. Cultural emergence of feature economy in an artificial whistled language. In E. Zee & W. Lee (eds.), *The 17th international congress of phonetic sciences: Hong Kong, China, Augustus 17–21, 2011: Congress proceedings,* 2066–2069. Hong Kong: City University of Hong Kong.

Walker, A. & J. Hay. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access. *Laboratory Phonology* 2(1). 219–237.

Walsh, M., B. Möbius, T. Wade & H. Schütze. 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. 2004a. *Self-organization and categorical behavior in phonology*. University of California Santa Cruz doctoral dissertation.

Wedel, A. 2004b. Category competition drives contrast maintenance within an exemplar-based production/perception loop. In J. Goldsmith & R. Wicentowski (eds.), *Proceedings of the seventh meeting of the ACL special interest group in computational phonology*, 1–10. Barcelona: Association for Computational Linguistics.

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. 2009. Variation, multi-level selection and conflicts between phonological and morphological regularities. In J. Blevins & J. Blevins (eds.), *Analogy in grammar: Form and acquisition*, 83–100. Oxford: Oxford University Press.

Wedel, A. 2011. Self-organization phonology. In M. Van Oostendorp, C. J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell companion to phonology*, 130–147. Oxford: Blackwell.

Wedel, A., S. Jackson & A. Kaplan. In press. Functional load and the lexicon: Evidence that syntactic category and frequency relationships in minimal lemma pairs predict the loss of phoneme contrasts. *Language and Speech*.

Whalen, D. H. 1981. Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary. *Journal of the Acoustical Society of America 69*. 275–282.

Whalen, D. H., A. S. Abramson, L. Lisker & M. Mody. 1993. F0 gives voicing information even with unambiguous voice onset times. *Journal of the Acoustical Society of America 47*. 36–49.

Wright, R. 2004. Factors of lexical competition in vowel articulation. In J. J. Local, R. Ogden & R. Temple (eds.), *Laboratory phonology*, *Volume 6*, 26–50. Cambridge: Cambridge University Press.

Wright, R. & S. Frisch. 2002. The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30. 139–162.

Yao, Y. 2009. *Effect of phonological neighborhood density on word duration in spontaneous speech*. Paper presented at the Fall Meeting of the Acoustical Society of America.

Appendix

Details of the computational model system

This computational system serves to illustrate the emergence and shift of groupings of tokens at two nested levels of organization, under biases in token-production and categorization. As in the body of the paper I will refer to these two levels as *segment* and *word* respectively. The word level of organization consists of four (or more) distinct word categories, each of which contains a list whose elements represent exemplars of previously encountered word tokens that were assigned to that category. Each element in the word category list is in turn a list of values representing exemplars of particular segment dimension categories, which can take values ranging from 0–100 (Figure A1). Position in the word-exemplar list is used to identify which segment-dimension the segment exemplar value maps to. In the models shown in the body of the paper, there are two distinct segment-dimensions, but the architecture can accommodate an arbitrary number. (To build a system that could accommodate gradient deletion or insertion of segments (or other elements), one would need to specify the particular segment dimension that a value maps to in some way other than by list position.) The position of every each word exemplar in the word-category list is a measure of its recency relative to other exemplars in that category, where the effect of recency on system behavior is modeled as an exponentially decaying activation within a category. At the beginning of each run, each word category is populated with a set of starting word exemplars.

The model uses two agents, each with an independent lexicon consisting of word categories containing word-exemplars as described above. Agents take

Word Category 1	Segment Dimension 1	Segment Dimension 2
Exemplar 1	12	23
Exemplar 2	11	20
Exemplar 3	14	22
Exemplar 4	13	24
. . .	. . .	. . .
Word Category 2	Segment Dimension 1	Segment Dimension 2
Exemplar 1	67	22
Exemplar 2	64	20
Exemplar 3	65	22
Exemplar 4	69	25
. . .	. . .	. . .

Table A1: The architecture of memory in the system, with example values.

turns producing an output from each of their word categories for each other, and likewise take turns categorizing and storing the outputs of the other. For the agent whose turn it is to produce a word, production begins with the random choice of a word exemplar from a category as an output target, where the probability of a particular choice is related to its relative recency in the exemplar list for that word category. In the model runs shown here, the activation of a exemplar is modeled as $e^{(.2j)}$, where j is its list position; this results in a exemplar at position 100 having an activation that is approximately .01 times that of an exemplar at position 1. The probability of an exemplar being chosen as a production target is its activation relative to the total activation of all exemplars in the category. Exemplars at list positions greater than 100 are discarded after every round to keep computation efficient; preserving more exemplars slows the rate of change in the system but otherwise does not qualitatively change system behavior.

Two biased versions of this target are then calculated independently of one another in terms of population vectors defined in relation to the original target segment values (Guenther and Gjaja 1996; Oudeyer 2002). At the word level, population vectors are calculated for the segment values in the target word relative to all segment values at the same position over all exemplars within that word category. At the segment level, population vectors are calculated for the segment values in the target word relative to all segment values on that dimension across the lexicon. To model the influence of both word and segment recency and similarity on production variation, the population vectors at each segment-dimension at each level are combined to create a new output that combines

information from both within-word category, and within-lexicon sources. The relative contribution of word versus segment population vectors to the output was fixed at .9.

The population vector with respect to a particular point within a particular segment dimension is a weighted average of all segment exemplars mapped to the category, where both the Euclidean distance from the target exemplar and activation influence each exemplar's contribution. This is conceptually the same as Nosofsky's Generalized Context model (Nosofsky 1988), modified to take exemplar activation into account. The formula used to incorporate these factors is given below, where p is the output population vector, y is each position within the segment dimension value of the target under production, w_y is the activation of the exemplar, x is the reference point chosen as the basis for production, and k is a scaling factor influencing the fall off of the contribution to the population vector of the point y relative to x:

$$p = \frac{\sum_y y w_y e^{-k|x-y|}}{\sum_y w_y e^{-k|x-y|}} \tag{1}$$

The value of k used in the simulations shown here is 0.2; a larger value of k reduces the effect of more distant values on the population vector.

Finally, a Gaussian random variable with a standard deviation of 3 is added to the output to introduce noise. This variable is biased slightly toward the center of the dimensional space, creating a fixed attractor at the center of each segment dimension in the system. The bias is calculated using a parabolic response curve given below, where b is the bias added to the output population vector, p is the output population vector, N is the number of points in the space and G is a constant; b is subtracted from outputs greater than $N/2$ (here, 50) and added to those below it.

$$b = \frac{(p - N/2)^2}{G} \tag{2}$$

The value of G used in these simulations was 5000, giving a bias toward the center of 0.5 at the edges of the continuum. All else being equal, this bias shifts the distributions of both categories toward the center of the dimension over time, i.e. toward 50, which corresponds to a simple model of articulatory undershoot (cf. Lindblom 1983; Pierrehumbert 2001). Note that because the anti-ambiguity effect modeled here (see below) acts to reduce the influence of ambiguous speaker outputs on the system rather than introduce hyperarticulated variants in response to potential ambiguity, this added random variable is the only source of new infor-

mation in the system which can prevent category collapse (cf. Figure 3 in the body of the text).

The listener compares the speaker's output to all of its word exemplars in each category, calculates a sum similarity score for each category using the Generalized Context Model for categorization (Nosofsky 1988), again modified to take activation into account as above in (1), where the scaling factor k is again .2. The speaker output is then stored as a new exemplar in the best fitting listener word category. Similarity bias operates in perception as well as production as evidenced in the well-known perceptual magnet effect (Kuhl 1991) and Ganong effect (Ganong 1980), and it can be modeled in this type of model using the same population vector strategy (Guenther and Gjaja 1996; Oudeyer 2002). However, the general behavior of this computational system is the same whether similarity bias and introduction of noise is modeled in once in production or perception or twice in both, so for simplicity I have only included it in production.

An anti-ambiguity bias is included at the level of the listener in this model in the form of a probability that a speaker output will not be stored as a new exemplar in the best-fitting category. This probability is the reciprocal of the similarity of that output to that category as calculated above, divided by the sum of its similarity to all categories. For example, if the similarity of a speaker output to the best fitting category is .9 relative to all categories, it will have a probability of .1 of not being stored.

After a speaker has produced an output target for each of its word categories, roles reverse.