Feedback and regularity in the lexicon[1]

Andrew Wedel
Department of Linguistics
University of Arizona
August 14, 2006

Abstract

Phonologies are characterized by striking regularity, from the stereotyped phonetic characteristics of particular allophones to the contextually conditioned alternations between them. Most models of grammar account for this regularity by hypothesizing that there is only a limited set of abstract symbols available for expressing underlying forms, and that an independent grammar algorithm predictably transforms these abstract symbol sequences into an output representation. However, this explanation for regularity is called into question by much recent research suggesting that the mental lexicon records rich phonetic detail that directly informs production. Given evidence for persistent biases favoring previously experienced forms at a number of levels of production and perception, I argue here that positive feedback within a richly detailed lexicon can serve as a source of regularity over many cycles of production and perception. Using simulation as a tool, I show that under the influence of positive feedback, gradient biases in usage can convert an initially gradient and variable distribution of lexical behaviors into a more categorical, and often simpler, pattern.

1.0 Introduction

Phonology concerns itself with two conceptually distinct domains of regularity, cross-linguistic and language internal. Many modern linguistic approaches have concerned themselves with cross-linguistic regularity – the observation that some structures and patterns tend to recur in many languages. In some (e.g., Trubetzkoy 1939, Jakobson et al. 1952, Archangeli and Pulleyblank 1994), this regularity stems from innate limitations on the alphabet of phonological features and their combinatorial possibilities, while others place limitations on the set of possible constraints governing phonological surface forms (Prince and Smolensky 1993). In other approaches, like Evolutionary Phonology (Blevins 2004, To appear a), a great number of crosslinguistic regularities are accounted for as byproducts of common pathways of language change.

However, *individual* languages do not recapitulate the cross-linguistic distribution of existing patterns and structures within their phonologies and lexicons, but instead exhibit some subset of these patterns consistently. For example, while neutralization of word-final obstruent voicing is a strong crosslinguistic tendency, given languages tend to either neutralize an obstruent voicing contrast word finally, or not. The sources of this kind of strong language internal regularity will be the subject of this paper. In generative models of phonological competence, regularity ultimately stems from strong limitations

[1] I would like to thank Adam Albright, Matt Gordon, Armin Mester, Joseph Salmons, the University of Arizona linguistics community, and expecially Juliette Blevins, Jaye Padgett and Adam Ussishkin for valuable discussion, comments and suggestions influencing the work in this paper. All errors remain mine.

in the kind and amount of information that can be stored and manipulated. However, a great deal of research over the last two decades indicates that speakers continually store and use a very broad range of detailed, redundant information in language production and perception (e.g., Hintzman 1986, Goldinger 1996, Johnson 1997, Alegre and Gordon 1999, Bybee 2002b). These findings, in turn, suggest a careful re-evaluation of the sources of regularity in phonological systems. In what follows I will argue that positive feedback loops engendered by this ongoing storage and reuse of information provide a plausible mechanism for the development of language-internal patterns of regularity.

In generative models of phonology and many of its precursors, regularity arises from the simplicity and abstractness of stored lexical representations and the operation of general algorithms that modify and enrich this encoded information in the process of production (for a review, see Kenstowicz 1994). In these models, the sound system of each language is maintained to be built upon a small set of abstract, contrastive features combinatorially arranged in a set of phonemes, which are in turn combined into ordered strings that form a set of sound-meaning units. These abstract lexical entries are informationally much simpler than corresponding speech outputs. The process of production therefore must involve mapping the simple, abstract features of a lexical representation into a more highly detailed motor representation (e.g., Levelt 1989). Despite this complexity, if the lexical representation that acts as a blueprint for the motor representation is built out of a limited set of abstract features, a consistent mapping from one to the other will necessarily result in a set of output forms that is highly stereotyped.

Languages are also characterized by contextually predictable correspondences between sounds in morphologically related forms. For example, in American English, the partially voiced coronal stop corresponding to the phoneme /d/ in the word *ride* is pronounced as a flap when followed by stressless vowel-initial suffix, as in the derived form *rider*. This is part of a larger general correlation between context and the realization of /d/, which is pronounced as a flap whenever it is between two vowels, the second of which is stressless. This pattern is nearly exceptionless, and is immediately extended to novel forms by native speakers. Languages are riddled with regular patterns like this, leading to the hypothesis in generative phonology that an additional stage exists between the lexicon and phonetic realization in which a set of algorithms pre-processes the lexical representation, adding, deleting, or changing information (see Kenstowicz 1994: chapters 2, 3). In rule systems, these algorithms take the form of re-write rules, while in constraint systems such as Optimality Theory (OT, Prince and Smolensky 1993), the algorithm takes the form of a system to adjudicate between competing constraints on form. To the extent that the featural patterns recognized by algorithms are found in many different lexical representations, the operation of the algorithm is an additional source of regularity in the set of output forms of a language. In summary, within the generative tradition, both rule and constraint models necessarily produce static regularity and regularity in alternation. This is because they (i) operate over informationally sparse lexical representations composed from a restricted set of abstract elements, and/or (ii) apply general algorithms that recognize a limited set of symbols to derive more complex output forms. The success of these approaches in modeling regularity in phonological behavior has been taken as support for the hypothesis that linguistic forms are strongly constrained by the alphabet of possible symbols and/or the capabilities of the grammar algorithm that produces outputs in a given language (e.g., Kenstowicz 1994: chapter 2, Hayes and

Steriade 2004). Below, I will use the term 'minimal memory model' to refer to theories like these that employ minimally detailed lexical entries or symbol sets.

1.1 Exemplars and rich memory models

A question rarely posed in modern generative work is why the language faculty would have evolved in this way. Why start from a minimal, abstract lexicon, and then obligatorily run lexical forms through a complex mapping algorithm to produce speech? Why not, for example, just memorize lexical forms and produce them directly, or memorize larger phonetic sequences shared by lexical forms and assemble these productively to produce speech, including novel forms? At least in part, the idea that information in the lexicon must be minimized appears to arise from the history of the field of informatics. At the time generative linguistic models were being developed in the 1950s-1960s, memory was thought to be limiting, as in the computers of the time. Limited memory required optimization of information storage through elimination of all extraneous or redundant detail (for discussion, see Anderson 1985: 134-139). As a consequence, models were developed that shifted as much work as possible from memory to a processor, which functioned to fill in predictable detail. As noted above, expansion of a small amount of information into a detailed representation by algorithm will necessarily produce stereotyped outputs. The observation that language outputs are highly stereotyped is consistent with this view (Kenstowicz 1994, Ch. 2).

However, as more research is done in areas as diverse as lexical access (reviewed in Luce and Pisoni 1998), word-specific phonetics (Pierrehumbert 2002), group dynamics (Mendoza-Denton to appear) and language change (Bybee 2001), evidence is rapidly accumulating that lexical representations are not solely abstract, but are richly detailed (reviewed in Goldinger 1996, Johnson 1997, Pierrehumbert 2002). In search of an adequate framework in which to explore these diverse findings, linguists of various stripes have turned to exemplar models.

Exemplar models of categorization propose that categories are composed of many detailed memories of instances, or *exemplars*, of that category, rather than, for example, a single prototype or a list of category features (for reviews, see Jacoby and Brooks 1984, Nosofsky 1988, Tenpenny 1995). As a result, categories can be populated with many differing exemplars of the 'same' thing – indeed, the only detail these exemplars must share is the fact of having been placed in the same category. For example, under an exemplar model the conceptual category 'bird' will contain many detailed sensory exemplars of actual birds, rather than a single, more abstract element or feature-list. Furthermore, because a percept can be stored as a new exemplar within a category, the contents of categories can evolve with experience.

Within exemplar models of language, a linguistic category contains many highly detailed exemplars of previously perceived members of that category. To model the greater influence of more recent memories, the activation level of exemplars is often modeled as slowly decaying over time (Nosofsky 1986, Pierrehumbert 2001a, Wedel in press). Production from a given category is generally modeled as activation of some exemplar (or local set of exemplars, Pierrehumbert 2002, Wedel in press), followed by mapping from that exemplar directly to a corresponding articulatory plan without the intervention of a grammar algorithm (Pierrehumbert 2001a, 2002; Wedel 2004, in press).

On the perception side, categorization proceeds by comparison of a percept to actual exemplars (Nosofsky 1986), or to generalizations derived from the range of exemplars within a category (Hintzman 1986; Goldinger 1996; reviewed in Tenpenny 1995). Upon identification of a percept with a category, a new exemplar is created in the corresponding category-space, or the activation of a previously stored, indistinguishable exemplar is raised (e.g., Kruschke 1992). As a consequence, each experience alters the entire category system slightly by changing the range and/or activation of component exemplars.

Experiments have shown that the narrow phonetic details of an utterance can be influenced by the details of recently perceived utterances (Tuller et al. 1994, Goldinger 1996, 2000, Guenther et al. 2004, discussed in Kelso (1995) pp. 207-212). This finding provides evidence for a production-perception feedback loop in adult speakers, in which non-contrastive phonetic details of what is perceived can be subsequently reflected in the details of what is produced (Pierrehumbert 2001a, Oudeyer 2002). The consequences of this production-perception feedback loop for phonological category evolution over time within a community of speakers have been explored in a number of recent papers. These studies suggest that a variety of phonological category change processes such as gradual word-specific lenition, category merger (Pierrehumbert 2001a, 2002, 2003), and contrast trading (Wedel in press) can be successfully modeled at a relatively fine-grained, mechanistic level with exemplar-based categories.

The ‘exemplar’ is a metaphor for a mechanism of information storage, and different models vary in their commitment to the details of this metaphor. For example, in models in which the exemplar is the sole unit of information storage, generalizations are proposed to be abstracted from the existing exemplar set on the fly as needed. Other models assume that exemplar-based memory serves instead as an intermediate buffer mediating between experience and a continually updating set of more abstract generalizations (reviewed in Tenpenny 1995; see e.g., Kuehne et al. 2000). Although exemplars form the sole unit of information storage in the simulations presented in this paper, the results are compatible with any similar model in which variation is recorded at some level within categories.

1.2 A rich-memory problem, a rich-memory solution

All fine and good, a minimal-lexicon proponent might say, but how can language internal regularity be explained in a rich memory model? Such models can account for a broad range of phonological behavior based on evidence for plentiful storage capacity in the brain, but once this storage capacity is assumed, what prevents rampant irregularity? A minimal-memory, generative model explains regularity with ease but has to do extra work to explain irregularity. On the other hand, a rich-memory model has the opposite problem: extra work is required to explain why irregularity in sound and word patterns is not the norm. And the fact remains that within a given language, some portion of the phonology and morphology is regular from the standpoint of linguistic analysis, and also productive, implying psychological regularity within the language user.

Here I argue that positive feedback within an evolving linguistic system can serve as a strong engine of the regularity that characterizes the form and distribution of words and sounds in a language. Within a rich-memory model of language production and perception, any factor that biases production targets or percepts towards previously encountered, similar structures creates a positive feedback loop increasing regularity over many cycles of production and perception (Pierrehumbert 2002, Oudeyer 2002, Wedel 2004). Below I show that in the context of similarity-promoting positive feedback, the rich detail of lexical and sound categories provides evolutionary pathways towards regularity and back away again, allowing transitions between relatively stable states over time.

Before modeling emergent regularity in this way, it will be helpful to define more precisely the term 'regular' in a rich-memory system where a wide range of variant pronunciations, performance errors, speaker identity characteristics, and other factors relevant to speech can be stored. A working distinction is often made among linguists between ‘phonological’ and ‘phonetic’ patterns: the former are categorical and consistent, and the latter can be gradient and variable. However, within a rich memory model in which a great deal of stored, possibly contradictory information participates in production and perception events, ‘categorical/consistent’ and ‘gradient/variable’ are more usefully thought of as conceptual endpoints on continua rather than being qualitatively distinct states. (For examples of phonological or morpho-phonological patterns treated in this way, see e.g., Skousen 1989, Albright and Hayes 2002, Ernestus and Baayen 2003, Wedel 2004.) In a rich memory system including a production/perception feedback loop, feedback pushes the future behavior of the system toward the categorical/consistent pole with regard to some dimension, while variation from any source pushes future behavior toward the gradient/variable pole. For the purposes of this paper then, the term ‘phonological pattern’ will be operationally used to refer to any static or relational pattern that holds of many lexical items with relatively high consistency. I will use the term ‘regularity’ in this context to describe a gradient property of consistency, where an exceptionless pattern is an endpoint on a continuum. The question explored here will be how feedback can promote greater regularity, and what factors may influence the patterns of regularity we find in a feedback-driven language system. In the process, I will argue that feedback can potentially contribute to a range of language-internal sound patterns that could be termed ‘regular’, such as patterns in context-dependence of allophones (section 3), and stress (section 4). I will use simulation as a tool to buttress these arguments.

Simulation is a useful approach for conducting experiments on the ramifications of particular system properties in a more controlled way than is often possible in the actual system of interest. While a simulation cannot prove that a system of interest really functions in a particular way, it can serve as an existence proof to show that a set of properties that the system is thought to have can interact in a particular way to produce a given type of result. Here I use simulations structured in parallel to psycholinguistically supported models of language to show that the ‘attractor’ of regularity set up by positive feedback between similar forms in production and perception creates categorical associations between features and structures qualitatively similar to those found in natural language. For simplicity, change in categorical associations over time is modeled here through error feedback within an idealized individual. Of course, real change of this sort is enabled not only through small incremental changes within an individual, but through the more substantial reanalyses that occur in language transmission (cf. CHANCE and

CHOICE within the Evolutionary Phonology program (Blevins 2004, to appear a); see also Kirby and Hurford 2002). However, phonological behavior can continue to shift over the lifetime of adults in response to changes in input (e.g., Sancier and Fowler 1997, Harrington et al. 2000), suggesting that some of the same mechanisms for linguistic change that operate during acquisition may continue to operate throughout adulthood. In the particular simulations presented here, the total number of exemplars stored by the system is kept purposefully low so that variant exemplars can have an appreciable influence on the evolution of the system within a convenient timescale. In that sense, the simulations can be thought of as modeling an individual permanently held in the early stages of acquisition where categories have not yet consolidated through long-term experience. Exploring additional patterns that may arise when more realistic cycles of acquisition over generations are modeled remains a task for the future.

Although the results that I present below are themselves consistent with models of language that do not invoke a grammar separate from the lexicon (e.g., MacWhinney 1998, Plaut and Kello 1999), they are also consistent with models in which a lexicon-independent grammar may exist, developing much of its form in acquisition in response to patterns in the learner's input and to emergent patterns in the learner's own production and perception (discussed in Pierrehumbert 2003). As a consequence, I will *not* be arguing that feedback-driven emergence of regular patterns through use itself tells us anything about the existence or non-existence of a modular grammar. I will rather be arguing that these findings should convince us that rich lexical representations, rather than being a problem for an account of phonological regularity, serve instead as a very plausible source of such regularity. Finally, I will argue that because self-reinforcing processes often begin tentatively and only occasionally snowball into categoricity, these results provide the beginnings of a mechanistic account of how initially local, word-associated phonological patterns can transition into neo-grammarian-style, lexicon-wide patterns (see also Bybee 2002a for similar arguments).

Section 2 provides background on positive feedback loops and how they can produce regular behavior in a rich memory system. In section 3, I introduce a simple, abstract model and via simulation, illustrate the emergence of patterns of regularity from conflicting pressures on form, including patterns that Optimality Theory attributes to the stipulated mechanism of strict constraint domination. In section 4, the workings of the model are illustrated more concretely by simulating the evolution of regular edge-aligned stress patterns. Finally, in section 5 I show that the model can account for the development of quantity-sensitive stress systems in which the heavy/light syllable weight distinction is correlated with the sonorant/obstruent ratio in the coda inventory (Gordon 1999, 2002).

2.0 Feedback and the emergence of structure

Within systems composed of many elements that interact with one another repeatedly over time, positive and negative feedback over these interactions can result in the development of unexpectedly complex, long-range order within the system (reviewed in Camazine et al. 2001) This process was discovered quasi-independently in many fields over the last century and has correspondingly many names, among them self-organization, auto-poieisis, and cybernetics. Positive feedback refers to a process that is auto-catalytic, that is, a process in which a given event makes a similar event more likely in the future. A simple example of this is the process of combustion, in which the reaction of a single oxygen molecule with a molecule of fuel produces heat that makes a subsequent reaction between nearby oxygen and fuel molecules more likely, and so on – producing the familiar, sudden burst of a struck match.

Feedback is negative if a given event makes a similar event *less* likely in the future. For example, limitations in food supply can produce negative feedback in reproducing populations, where creation of a single additional offspring reduces the probability that *any* offspring will survive to reproduce. In many biological population systems, the conflicting interaction of the positive feedback from reproduction and the negative feedback from resource limitation results in a steady state population density at which just the number of offspring survive to maintain the current population density. In systems in which the effect of resource limitation on reproduction is temporally delayed, complex oscillations in population and resource density can result[2].

Many complex patterns in nature are either known or hypothesized to derive from positive and negative feedback interactions over time (Kaufmann 1995, Camazine et al. 2001). Some familiar kinds of patterns include the characteristic shape of thunderstorms, patterns of form in shells and flowers, the social organization of insect colonies, the schooling behavior of fish, portions of the human immune system, stock market trading patterns, and so on. Feedback-driven structure is found in so many disparate kinds of systems because the requirements for feedback-driven structure formation are themselves quite simple. In general, feedback-driven accumulation of long-range structure requires that a system contain repeatedly interacting elements, where interactions between elements are non-identical, and where the effect of a particular interaction persists in the system long enough for it to influence the rate or character of future interactions – that is to say, long enough for it to provide feedback. This set of requirements is simple enough that it is likely to be met at some level by many different kinds of systems and as such, it would be very surprising indeed if some kinds of language structure were not found to be influenced by feedback mechanisms (Lindblom et al. 1984, Cziko 1995).

2.1 Positive feedback in language

A growing body of work ranging over the evolution of semantics (e.g., Oliphant 2002, Steels and Kaplan 2002, Kirby and Hurford 2002, Smith 2005), semantics-syntax mappings (e.g., Steels 1998, Batali 2002, Brighton et al, 2005), morphology (e.g., Hare and Elman 1995, Batali 2002, Bybee 2002a), syllable structure (Redford et al. 2001) and vowel-systems (e.g., Lindblom et al. 1984, Joanisse and Seidenberg, 1997, de Boer 2001) suggests that feedback cycles within language communities can contribute to accounts of a broad range of linguistic patterns. At the same time, work within a generative framework has also provided evidence that a similarity attraction effect plays a role in generating patterns in the lexicon (e.g., Burzio 2005). In this section I provide a brief overview of several mechanistic routes for positive feedback in phonological categories

---

[2] NetLogo (http://ccl.northwestern.edu/netlogo/) provides a wide range of educational, user-directed online simulations of feedback-driven processes. For an illuminating simulation of an oscillating predator-prey system, point your browser to http://ccl.northwestern.edu/netlogo/models/WolfSheepPredation.

that increase local similarity, using three linguistic phenomena as examples: (i) the observed coherence of sound categories, (ii) sharing of features between sound categories, and (iii) segment substitution within words.

At the level of non-contrastive phonetic detail, at least two general mechanisms have been proposed that independently serve to create attractive biases toward frequently encountered variants: motor entrenchment in production, and the magnet effect in perception. Motor entrenchment is a general property of motor systems in which practiced routines form attractors which bias future motor execution in relation to similarity (Zanone and Kelso 1992, 1997, see also Saltzman and Munhall 1989, Kelso 1995). As a consequence, for a given segment or segment sequence, the most frequent motor routine variants should be steadily reinforced relative to less common variants (Bybee 2002a). This is relevant for the current discussion because, as pointed out earlier, a conceptual issue with a rich-memory model of the lexicon is that it, in principle, allows nearly unlimited variation. If all variant pronunciations of a given sound are stored and then potentially reproduced, a sound category should inexorably broaden over time as it collects variants, and then variants of variants and so on (for further discussion, see Pierrehumbert 2002). However, sound categories do not broaden over an individual's lifetime; rather, they become steadily more coherent through childhood (e.g., Barton 1980, Lee et al. 1999). Motor entrenchment provides a mechanism to counterbalance the steady accumulation of variation, promoting a steady state level of category coherence.

On the perceptual side, the perceptual magnet effect (Kuhl 1991, 1995) provides another source of positive feedback bias which can act to promote category coherence. The perceptual magnet effect refers to the finding that the perceptual space is warped toward category centers relative to objective stimuli dimensions, such that percepts near category centers are perceived as closer together than they actually are, while percepts near category boundaries are perceived as farther apart. The result is that percepts tend to be biased systematically toward the centers of categories relative to the stimuli that gave rise to them. Within models of the perceptual magnet effect in which warping precedes categorization (Guenther and Gjaja 1996, cf. Kuhl 1995, Lacerda 1995), this systematic warping pulls similar pronunciations closer together over time through feedback between perception and production (Wedel in press).

Just as distinct motor routines *within* a sound category influence one another, motor theory suggests that motor routines belonging to distinct, but similar segment and segment sequence categories should influence one another to become more similar over time. Consistent with this, recent evidence suggests that gradient, subsegmental gestural errors creating variously 'hybrid' segments do frequently occur (Pouplier et al. 1999, Wright and Frisch 2002, Pouplier and Hardcastle 2005). If such a hybrid variant is stored as a motor exemplar for a given sound category, it brings the average of that category closer to the other category contributing to the hybrid. Recall that within exemplar models with a production/perception feedback loop, an exemplar stored in a category contributes both to future production events from that category, and to identification of percepts with that category. A resulting prediction of these models is that hybrid variants should serve as catalysts providing a pathway for increased sharing of gestures between similar sounds, up through and including full category merger (see Hansson, this volume).

Finally, at the level of the segment as an abstract category, full segmental substitution errors are more likely to occur between similar than dissimilar segments

(Shattuck-Hufnagel and Klatt 1979, Frisch 1996), and appear to be more likely the more frequent the resulting segment sequence (Dell et al. 2000). Long-term feedback over this bias would be expected to produce a lexicon in which licit properties and sequences are not randomly distributed, but are concentrated in islands of local regularity (see e.g., Ernestus and Baayen 2003)[3].

Recording a single variant is unlikely to substanstially change the character of a category, but in the context of positive feedback between production and perception, all of the error biases discussed above create a persistent asymmetry in the distribution of variant exemplars that are stored. Most of the time, these variants will remain on the fringe of the category, but because every stored exemplar makes production and processing of a similar form more efficient in the future, any variant can occasionally serve as the seed of more substantive category change, particularly during language acquisition.

2.2 Similarity-based error biases and the emergence of regularity

In many models of language change, error[4] provides a significant proportion of the variation that makes up the raw material for reanalysis (e.g., Ohala 1981, 1989, Labov 1994, Blevins 2004). Here I will argue that error biases that favor previously encountered forms such as those discussed above provide a mechanism for development and entrenchment of language internal regular patterns. Regularity arises in this situation because under positive feedback, gradient behavior is unstable over time. To see why this is true, consider a simple simulation starting with a field of randomly distributed squares of two different shades (Figure 1a). In each round of the simulation, each square checks the shades of the squares immediately surrounding it, and if the majority shade is different, it changes shade to match; if the distribution is equal it has even chances of changing or staying the same. Figure 1b illustrates the distribution of the two shades after several rounds. Note that the distribution has become strikingly more regular, as squares progressively change to match their local neighborhood. As each round passes, curves and bends in the distribution are straightened out in favor of the locally more densely populated shade, producing progressively larger uniform areas and correspondingly smaller areas of contact (Figure 1c). The steady reduction of area of contact is driven by the fact that squares at a shade boundary are likely to switch back and forth, while those surrounded by uniform shade do not change. As a boundary shifts back and forth randomly, any change that places a given square in a majority one-shade neighborhood

[3] Many morphological changes are also consistent with a feedback process that increases local, as opposed to global regularity (Joseph and Janda 1988, 1997). This can be seen in certain kinds of paradigm leveling, in which forms within a paradigm become more consistent with one another in some way (Bybee 1985, Hock 1991, Albright 2002). Likewise, regularly inflecting forms are often produced irregularly in parallel to a phonologically similar, irregularly inflecting form (Long and Almor 2000, Almor in press), and are judged better than irregularly inflected regular forms that lack a phonologically similar irregular neighbor (Albright 2002).

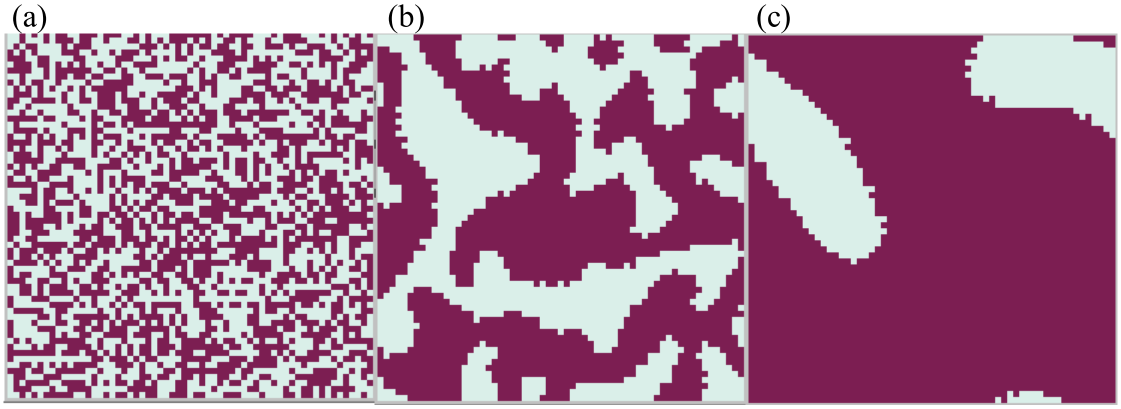
[4] The term 'error' will be used here in the loose sense of variation in production or perception from some usefully defined norm caused by noise at some level in the system.

will be relatively more stable, minimizing boundaries over time[5]. This is an example of how a bias toward local similarity can result in global regularity given sufficient time.

Returning to language, in a rich-memory system in which experience leaves a lasting trace that can influence subsequent production and perception, error biases toward similarity at some level of categorization should promote the development of crisp boundaries in behavior at that level of categorization (Wedel 2004). In the next section, I introduce a simple heuristic model of a production/perception loop that demonstrates the development of regularity under positive feedback, and then turn to more complex regular patterns that arise when multiple contextual biases compete.

Figure 1

(a) (b) (c)



3.0 Modeling positive and negative feedback in the lexicon

A single phoneme category is often made up of multiple relatively distinct, context-dependent allophones[6]. A range of laboratory results suggest that for many allophones, errors in production or perception are relatively reduced for that allophone in its conditioning context (e.g., Ohala 1983, 1990, 2005). Within a framework for language change that allows for incremental, error-driven modification of categories (e.g., Evolutionary Phonology, Blevins 2004), the accumulation of more stable versions of a sound in a particular context should promote the development of context-conditioned variants within a sound category. For example, many languages exhibit two relatively distinct, context-dependent allophones of /l/, velarized and plain, which differ among other things in the backness of the tongue body (Ladefoged 2006). In Georgian for example, the choice of plain [l] is conditioned by a following front vowel (Tschenkeli

---

[5] This is conceptually parallel to the separation of oil and water into volumes with minimum contact driven by water molecules' greater affinity for each other (Wedel 2004, 41-42).

[6] How distinct, and how context dependent is an ongoing research topic (see e.g., Sproat and Fujimura 1993). The general model discussed here does not assume that a given physical instantiation of a sound category will always be clearly assignable to an allophone category or will be absolutely correlated with a given context. Rather, in consonance with exemplar models in general, production is modeled as a stochastic process that is influenced by many factors, resulting in overall distributions that may be significantly consistent at a given level of granularity.

1958). Under this model, the Georgian pattern suggests the hypothesis that plain [l] is more stable in the environment of a following front vowel than velarized [l].

If it turns out to be true that the plain [l] is more stable in this context, shouldn't all languages have plain variants of /l/ associated with following front vowels? How do we explain the observation that there exist languages that seem to only use one [l] variant in *all* contexts, such as dialects of English that employ only a relatively velarized [l] (Carter 1999, Johnston 2002), and others that employ only a relatively plain [l] (Hickey 2004)? More puzzling yet, how could we explain the fact that there are other dialects of English that exhibit both allophones of /l/, but conditioned by an entirely different context, namely syllable position (Ladefoged 2006, p. 76)? This is not an isolated problem, of course. A basic finding of phonology is that allophones are regularly associated with some context, specific or default, and that this context can be different in different languages.

Optimality Theory accounts for the language-specific suppression or permission of allophonic variation through language-specific ranking of faithfulness and markedness constraints. However, I will argue that a quite different, evolutionary model can explain this crosslinguistic variation in the patterns of regularity in a more straightforward way. Specifically, I will show here that competition between two kinds of error in the context of positive feedback allows phonological systems to potentially occupy a variety of distinct, stable states. In this model, error at the level of the phonological category in favor of its more frequent variants promotes category uniformity, while context-sensitive error in articulation/perception promotes the development of as many variants as there are biasing contexts. Because positive feedback pushes the system toward global consistency, competition between these two effects results in a tendency for the system to settle on a restricted set of the possible categorical patterns, ranging from the case of a single allophone for a phonological category, to a relatively limited set of regularly conditioned variant allophones. Below, I describe a set of simple simulations that illustrate the competition between these two kinds of error and the categorical patterns that result.

The model of pattern change for the simulations in this paper is built on a simple loop between production and perception (Figure 2). Because we are not modeling contrast effects that arise in categorization of percepts, the simulation can be simplified to model just one lexicon interacting with itself[7]. For the simulations in this section, the 'lexicon' consists of entries built from ordered segment categories taken from the set {a, b, c, ..., x}. Each of these categories will be treated as containing one average exemplar, with the exception of the category /x/, which will be idealized as a distribution having two stable sub-categories, or 'allophones' represented by the values '1' or '-1'. In a more complete model, both larger (e.g., segment sequence) and smaller (e.g., allophone) categories would be included, each consisting of a distribution of exemplars[8]. However, the goal here is not verisimilitude, but rather to illustrate, as clearly as possible, how feedback loops can operate in a rich-memory system to produce regularity. Therefore, in order to keep the pathway of information flow in the simulation as transparent as possible

[7] For simulations of contrast maintenance in a similar model requiring a multiple-speaker system, see Wedel 2004, in press.
[8] For examples of simulations of the evolution of exemplar-based sound categories, see Pierrehumbert 2001a, 2002; Wedel 2004, in press.

I only model category substitution errors here and show how these kinds of errors alone can produce patterns of interest.

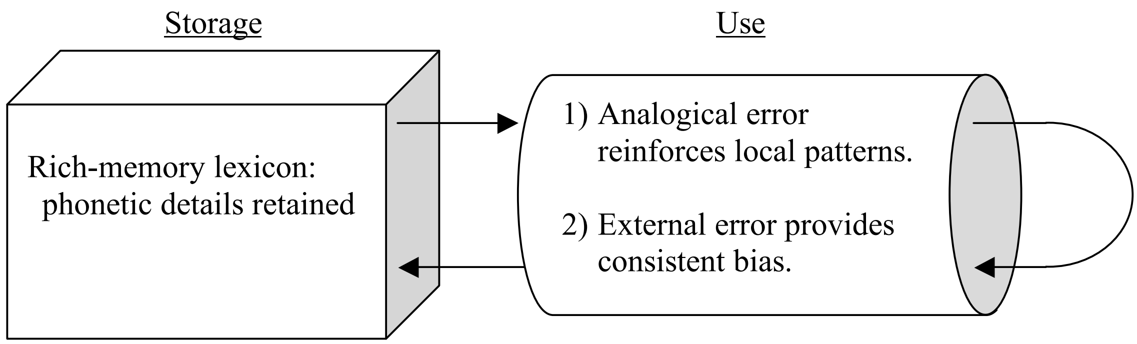
The production/perception loop proceeds in each cycle by retrieving each entry and restoring it, replacing the original to mimic the slow replacement of old exemplars with newer ones. Changes in form arise when error intervenes in this loop, changing the value of /x/ in a produced form before it is restored. This restorage of errorfully produced forms – each of which will go on to serve as a model for production in the next cycle – results in error feedback between cycles. As we saw above, positive feedback in and of itself promotes the development of categorical patterns. Here, we will be interested in the regularity relationship between individual lexical entries and the lexicon as a whole rather than the regularity relationship between individual category exemplars within an entry and the lexical entry as a whole (cf. Wedel in press). As a consequence, the computational shortcut of imposing quantal values on /x/ rather than a distribution of values does not change the relationship that we are modeling at this level of abstraction.

In this model, linguistically relevant patterns arise through the interaction of two conceptually different sources of variation, which I will refer to as analogical error and external error. Analogical error is a cover-term for error biased toward common patterns in experience of the sorts described in section 2.1 above. Analogical error refers then to error patterns that are derived from details of development of the system in context, and as such is contingent on the particular history of the larger system. In the simulations presented here, this error is modeled as a bias towards replicating previously encountered associations between the value of /x/ and the category labels /a, b, x/ in proportion to their relative type frequency, as recorded in the current lexicon. Frequency- and similarity-dependent analogical error creates positive feedback because the more such errors reinforce a given association in the lexicon, the stronger the error tendency toward this association becomes. The influence of both similarity and type frequency on a number of language phenomena has been well established; for discussion in a variety of contexts see e.g., Skousen 1989, Albright and Hayes 2002, Bybee 2002a, Ernestus and Baayen 2003, Almor in press. The algorithm for calculating and applying this error within the simulations presented here is described in the appendix.

In contrast, the term external error covers less experience-dependent biases toward changing one set of category values toward another. This error is modeled after the kinds of quasi-language-independent factors of physiology, perception and processing that have been proposed to underlie many cross-linguistic regularities (e.g., Ohala 1989, Lindblom 1998; see Hayes and Steriade 2004, Blevins 2004 for reviews). This error is held constant within a simulation, and is uninfluenced by patterns in the lexicon (cf. CHANGE in the framework of Evolutionary Phonology, Blevins 2004). Of course, it is an oversimplification to split error into these two opposing categories, as experience and context are likely to modulate the likelihood of errors of all types. For example, frequency of use impacts both motor and perceptual accuracy, as discussed in Bybee 2001. However, for simplicity, no such interaction is modeled in these simulations.

This model has been intentionally kept very schematic for the following reason. As will be seen below, even in simulations of a feedback-driven model with as few 'moving parts' as this one, relatively complex patterns can arise. Keeping the model maximally simple and transparent makes it easier to identify the causal relationships within the cycle that underlie the development of these patterns.

Figure 2. Pattern of information flow within the model

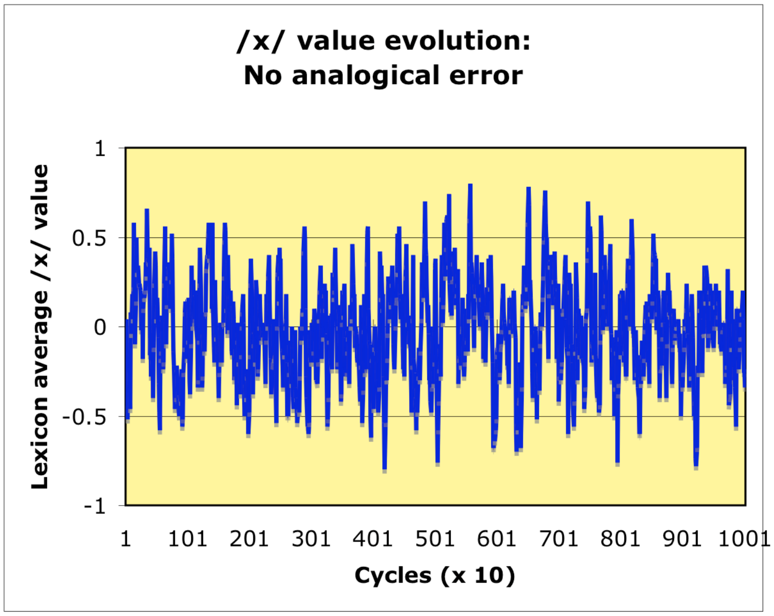


3.1 Analogical error promotes uniformity

To illustrate the contribution of analogical error to evolution of the lexicon in this model, let us start by comparing two simulations: a control simulation lacking both analogical and external error, and a simulation that includes analogical error alone. Both simulations include a further random error set at a 1% chance that an /x/ value will flip in any production event. These simulations follow the evolution of a lexicon with 100 lexical entries each containing the category /x/ and a randomly assigned additional category taken from the set {d, e, f, ..., w} which functions to supply additional random similarities and differences between words. At the start, the lexicon is balanced with 50 entries containing [x: +1] and 50 containing [x: -1]. Because the model presented here abstracts away from contrast maintenance to focus on feedback mechanisms promoting development of analogical patterns, no cost is imposed if lexical entries happen to be identical (cf. Wedel 2004, in press, for simulations of contrast maintenance in a feedback driven model).

Figure 3 shows the evolution of the average value of /x/ over the entire lexicon for 10,000 cycles in the presence of random error alone. Since the two values for /x/ are coded as [+1] and [-1], a lexicon with equal numbers of each value will show an average /x/ value of zero, while a lexicon with a tilt toward one or the other value will show an average closer to 1 or -1, respectively. Note that the lexicon could only rarely ever show an average /x/ value of exactly 1 or -1, because some lexical items in any round are likely to be produced with the minority /x/ value.

Figure 3



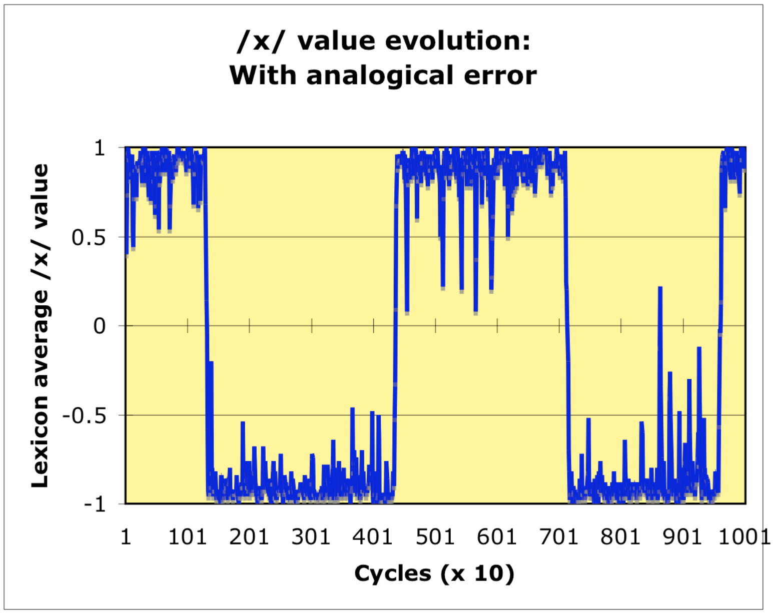
In the control simulation lacking analogical error shown in Figure 3, the mean /x/ value of the lexicon as a whole at any time remains close to 0, as we expect since every lexical entry is fully independent of every other. When many variables (in this case the values of /x/ in the lexicon) are randomly distributed, the central limit theorem states that the sum will be approximately normally distributed. As a consequence, the probability that the mean value of /x/ will approach the limits of +1 or -1 is very low[9]. This means that within this model, without the action of analogical error the distribution of feature values in the lexicon tends to maximal *non*-regularity.

The simulation illustrated in Figure 4 is the same, except for the inclusion of analogical error. Here we see a strikingly different pattern, as the value of /x/ veers immediately to an extreme and stays there, switching only occasionally to the other extreme over the course of the simulation. This pull to extremes occurs because when a particular /x/ value attains a sufficient majority in the lexicon through random chance, analogical error favoring that value begins to become ever more probable, pulling the entire lexicon further in the same direction. This is conceptually the same as the process we saw above in Figure 1, where gradience (there in the form of a uniform distribution of shades across the field of squares) is rendered unstable by positive feedback. Within the simulation in Figure 4 however, continual external error at a rate of 1% per lexical entry per cycle occasionally injects enough of the alternate /x/ value into the lexicon to weaken the direction of analogical error and allow a switch to the alternate value. Note that the

[9]There are many different ways that /x/ values can be distributed in the lexicon to achieve an average near 0, but there are many fewer distributions that can result in an average near +1 or -1. For example, there is only one set of /x/ values that provides a lexicon-average of exactly +1, namely the one in which every single /x/ value is [+1]. An analogous situation occurs with the range of possible outputs of a throw of two dice. There is only one way to roll a two or a twelve: two ones, or two sixes respectively. However, there are two ways to roll a three, three ways to roll a four, and the most common outcome is seven at the middle of the distribution with six distinct ways: {(1, 6); (6, 1); (2, 5); (5, 2); (3, 4); (4, 3)}.

push toward categorical behavior comes entirely from positive feedback in this system – there is no extrinsic bias in favor of either of the two stable states in this system.

Figure 4



3.2 Bias conflict and pattern-leveling

Consider the crosslinguistic distribution of velarized and plain /l/ mentioned earlier. There are some languages with only velarized or plain lateral approximants in all contexts. And there are other languages where velarized and plain laterals are variants or 'allophones' of the same sound category or 'phoneme'. In some languages, one variant is associated with a particular phonological context: a plain lateral may be restricted to occuring before front vowels, as in Georgian; or plain laterals may occur only in syllable onset positions, as in some dialects of English. If one assumes that the context-influenced pathways for development of plain vs. velarized laterals are present, to some extent, in most languages which have a voiced lateral approximant phone, then two questions immediately present themselves. First, why aren't distinctly velarized and plain [l] present in all languages that have /l/? And second, how does a pattern arise associating one of these variants with one possible conditioning context to the exclusion of other conditioning contexts? For example, how can it be that the distinction in velarization of /l/ in English is primarily associated with syllable position, while in Georgian, it is dependent on the backness of a following vowel? In this section we will see how the pattern-smoothing effect of analogical error competes with the pattern diversifying effect of external, contextual biases, resulting in a limitation of the number of factors that can condition variation. In the context of positive feedback, this competition can produce a system in which both uniform category exponence and contextually conditioned allophonic variation are stable states. As a result, lexicons under the very same starting conditions can settle unpredictably into one or the other kind of state, depending on small random variations in the initial path taken.

In the simulation shown in Figure 4, two categorical patterns developed solely through positive feedback, without any extrinsic bias toward one /x/ value or the other. To model the interaction of distinct contextual biases, we set all lexical entries in the simulation to include either the categories /a/ or /b/, and introduce external contextual biases in error for the value of /x/ based on the presence of [a] or [b][10]. Specifically, the presence of [a] increases the relative probability of error toward [x: +1] to 3%, and the presence of [b] increases the relative probability of error toward [x: -1] to 3%. The context free random probability of an /x/ value changing remains at 1%. The lexicon consists of 100 entries, 50 of which contain /a, x/, and 50 /b, x/, where the starting value of /x/ is balanced across both sets of entries. Again, an additional random category is appended to each lexical entry, here taken from the set {c, d, e...w}. Recall that analogical error can be created by a significant association of /x/ values with any of the category labels included in the simulation, including /a/ and /b/ as well as the added random category. There are four possible major patterns of association that are available as attractors for the system solely through positive feedback: association of the value of /x/ with the category label /x/ itself, resulting in consistent /x/ values of either [+1] or [-1] across the lexicon; or association of a particular value of /x/ with /a/, and the other with /b/ or vice versa. Minor patterns are in principle possible in the association of a particular value of /x/ with the randomly assigned category label in each lexical entry.

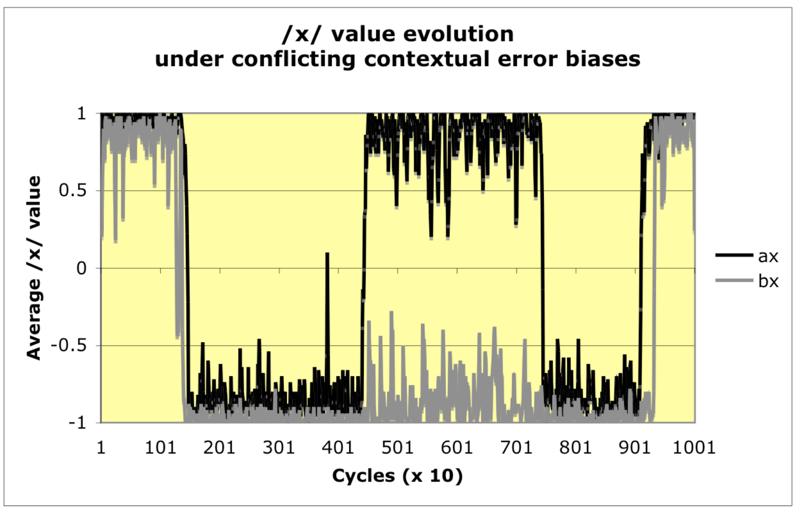
Consistent, divergent error biases for /x/ values conditioned by [a] and [b] encourage the development of a lexicon in which [x: +1] is associated with /a/ and [x: -1] associated with /b/. However, one might also expect global positive feedback between the value of /x/ and its own label /x/ to be able to override these more local patterns and enforce one of the two possible lexicon-wide values of /x/. Looking closely at Figure 5, we can see that in fact these three patterns emerge. In the initial portion of the simulation through approximately cycle 1000, the average value of /x/ for the entire lexicon is near +1, despite the fact that low level error in production of outputs containing /b/ promotes an /x/ value of [-1] in those outputs. Shortly after 1000 cycles, the value of /x/ flips for the entire lexicon toward -1, and remains there for another 3000 cycles. Periods such as these, in which /x/ values are consistent across the lexicon, are stabilized by analogical associations between all category labels with one value of /x/ that overrides persistent biases toward association of a particular /x/ value with /a/ versus /b/.

Near cycle 4500 however, persistent error resulting in /x/ values of [+1] in outputs containing [a] created a sufficient nucleus of /a, x: +1/ entries in the lexicon to support a separate stabilizing association between /a/ and /x: +1/, with the result that /x/ values for lexical entries containing /a/ jumped up to [+1], leaving those containing /b/ behind. Here, the associations of the category label /x/ with its possible values [+1] and [-1] have become equivalently strong, resulting in no net analogical error bias from this source. The development of the two ‘allophones’ of /x/ here is initially prompted by external contextual biases, but once in place, the stable pattern is maintained through their consistent associations with /a/ and /b/ within the lexicon. The three stable states that appear in Figure 5 are parallel to the distribution of allophones of /l/ in language, where we find languages with one [l] variant that remains relatively constant across contexts, and others that have developed two conditioned variants in association with some context.

[10] Slash brackets are used to refer to category labels, and square brackets to output values.

Optimality Theory could account for the various output patterns in Figure 5 through the relative ranking of context-free constraints violated by either value of /x/, and context-sensitive constraints violated by particular values of /x/ in conjunction with [a] or [b]. For example, the uniform output value of [x: +1] in the 1-1000 cycle range could be accounted for by ranking a constraint against [x: -1] above a constraint against [x: +1], while the opposite pattern found in the 2000-4000 cycle range could be accounted for by the opposite ranking. In turn, the pattern of context dependent variation in /x/ values that emerges in the 5000-7000 cycle range could be accounted for by ranking the two context-sensitive constraints above the context-free ones. The feature of Optimality Theory that allows these conflicting categorical patterns to emerge from the same general system is the stipulated property of constraint dominance: all of these constraints may be present in the grammar, but in the event of constraint conflict, the higher ranked constraint wholly determines the outcome. In contrast, in the model simulated here pattern 'dominance' arises naturally under the well-supported assumption that production and perception are subject to similarity-based positive feedback. In the next section, we will discuss this feature of the model further in the context of the claim that phonological patterns are often more coarsely grained than the phonetic trends that give rise to them.

Figure 5



3.3 The coarseness of the grammar

It has been noted that statistical regularities in the speech stream seem to be more complex and fine-grained than the phonological generalizations that develop from them (Hyman 1977, Gordon 2002, Pierrehumbert 2001b, 2003). In a rich-memory model, analogical error minimizes diversity of behavior along some dimension over time, while any sources of variation promote greater diversity in behavior. Within this model, the competition between these two general tendencies results in the stabilization of phonological patterns that are on average more coarsely grained than would arise from a straightforward mapping of all external biases onto the lexicon. For example, in the simulation shown in Figure 5 above, the evolution of a category under the influence of two opposing context-specific biases often still results in relatively uniform category exponence, because positive feedback favors uniformity. In this section, we will see that analogical bias can transform a more complex set of interacting external biases into simpler patterns that would be accounted for in Optimality Theory through the mechanism of strict constraint domination.

As discussed in the introduction, generative models of phonological competence favor coarse-grained phonological patterns through limitations on the alphabet of underlying symbols and their combinatorial possibilities, and/or on the rules or constraints that govern output forms. Much of Optimality Theory's ability to account for the coarse-grained nature of grammatical patterns lies in its claim that there is a limited set of universal constraints, and that there is a limited mechanism for their interaction, in particular that the choice of optimal outputs proceeds through satisfaction of constraints in ranked order.

Standard Optimality Theory (Prince and Smolensky 1993) stipulates that the choice of optimal outputs proceeds through satisfaction of constraints in ranked order. The principle of *strict domination* further specifies that ranking is absolute: no degree of potential violation of lower ranked constraints can ever compel violation of a higher ranked constraint. This principle functions to limit the complexity of the patterns that can be generated with an Optimality Theoretic grammar. Strict domination can often be found paraphrased informally as, 'Lower ranked constraints can't gang up against a higher ranked constraint' but another informal restatement that will be especially useful here can be given as "The outcome of multiple constraint conflicts reproduces the outcomes of the component pairwise constraint conflicts". This way of stating the principle makes it clear that its operation keeps phonological patterns simpler than they would be otherwise. Further, because the set of forms in which two particular constraints conflict must be equal or larger than the set of forms in which those two plus another constraint all conflict, the principle of strict domination can be seen as a prediction that winning patterns are associated with typologically more frequent constraint conflicts.

This limitation allows Optimality Theory equipped with a particular finite set of constraints to predict that certain complex patterns cannot exist. For the purpose of illustration, consider a proposed markedness constraint banning breathy-voiced vowels (abbreviated *NON-MODAL, Gordon 1998), and another banning low, round vowels (abbreviated *ROLO, Kaun 1995). These markedness constraints can be ranked with faithfulness constraints preserving vowel features (here lumped together as FAITH), to produce a factorial typology of possible grammars as regards their tolerance for breathy voicing of vowels and round, low vowels:

Table 1. Factorial Typology

Factorial Typology: FAITH X *ROLO, *NON-MODAL

	*NON-MODAL >> FAITH	FAITH >> *NON-MODAL
*ROLO >> FAITH	no breathy voiced vowels no low, round vowels	breathy vowels OK no low, round vowels
FAITH >> *ROLO	no breathy vowels low, round vowels OK	breathy vowels OK low, round vowels OK

This typology shows us that this set of constraints can generate languages that allow or disallow breathy vowels and low, round vowels respectively. However, it cannot generate a language that allows vowels that are breathy *or* low-round, but draws the line at vowels that are *both* breathy, *and* low and round. Generating such a language would require a new constraint specifically banning the conjunction of these properties[11,12]. The principle of strict domination functions to prevent the additive interaction between simpler constraints within a grammar, and thereby limits the particularity of the lexical patterns generated by a given ranking.

However, while strict domination allows OT to accurately describe many phonological systems, it sits uneasily with the notion, increasingly well-represented within the field, that constraints on grammatical patterns are directly or indirectly related to phonetic biases based in articulatory and perceptual difficulty (e.g., Donegan and Stampe 1979, Archangeli and Pulleyblank 1994, Steriade 1997, 1999, Blevins and Garret 1998, Hayes 1999, Hayes and Steriade 2004, Blevins 2004, Blevins to appear a). This unease arises because it is difficult to see how biases with sources outside the grammar, such as physiological constraints on articulation or perception, would not interact additively in some overall performance cost. For example, continuing our example from above, if breathy phonation is more articulatorily difficult than full voicing (reviewed in Gordon 1998), and a low-round vowel is more articulatorily difficult than a high-round vowel (Kaun 1995), then these costs should compound at some level: a low-round, breathy vowel should be harder *in toto* than either a high-round breathy vowel or a low-round modally-voiced vowel. Within Optimality Theory however, the inability of low ranked constraints to gang up on higher ranked constraints amounts to a statement that though costs may sum at the level of articulation and perception, they cannot at the level of phonology unless a specific constraint exists that by itself penalizes that combination of costs (as in, for example, a constraint that conjoins a penalty for voicing in obstruents with a penalty for codas to account for syllable-final obstruent devoicing (Ito and Mester 2003)). Any language pattern that suggests that some particular costs are behaving additively represents evidence for a constraint that penalizes the conjunction of those costs. The fact that constraint conjunction has frequently been invoked in Optimality Theoretic work suggests that strict domination does not serve as an accurate mechanism to limit the power of Optimality Theory. It is too strong because independent constraints can show additive effects when operative within particular domains (e.g., Ito and Mester 2003). And when it is weakened by the allowance of constraint conjunction to allow these kinds of additive effects, there is significant overgeneration of unattested systems (discussed in Padgett 2002).

Here, we will see that when the interaction of multiple external error biases is simulated within our simple model, regular patterns can evolve that are consistent with the strict domination principle of OT, despite the fact that these interactions do compound in production, as we expect they should. In particular, when multiple patterns potentially conflict in a single output form, we will see that pattern-smoothing from analogical error

---

[11] Admitting a new constraint that bans breathy, round, low vowels is equivalent to conjoining the two markedness constraints *RoLo and *Non-Modal (Ito and Mester (2003)).

[12] Such a constraint might seem odd to a phonologist. However, if phonological pattern development is predicted to be influenced directly or indirectly by difficulty, any combination of features within some domain that incurs greater effort should be in principle subject to a constraint banning that combination.

promotes outcomes that follow the individual pairwise outcomes of pattern conflict, just as predicted by the principle of strict domination. Interestingly, we will see that patterns can also arise that violate strict domination (i.e., would require the addition of a new constraint), but only when violations of multiple external biases are associated in a sequence that is typologically frequent in the lexicon.

Here, we will consider the interaction of three external biases: 1) A context-free 5% error bias toward [x: +1] in production; 2) A 3% bias toward [x: -1] conditioned by the presence of [a]; and 3) A 3% bias toward [x: -1] conditioned by the presence of [b]. These error biases are summarized below.

1) {x} → [x: 1] at 5%
2) {a, x} → [x: -1] at 3%
3) {b, x} → [x: -1] at 3%

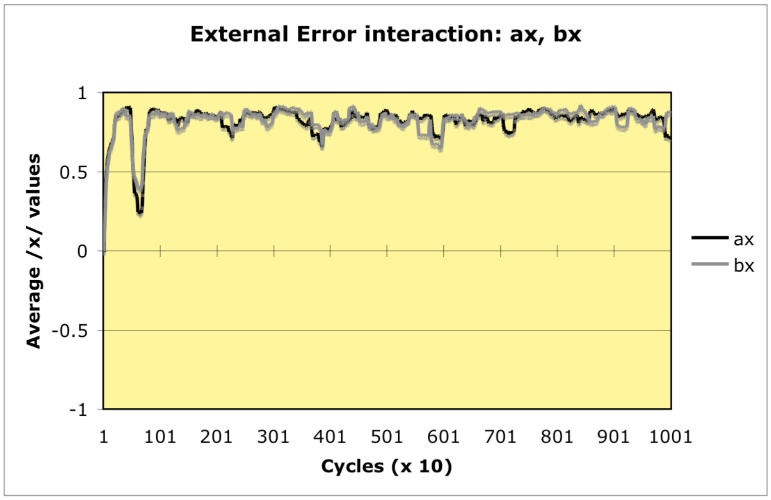
Within the simulation, these biases are modeled as independent of one another, as we expect for biases with distinct sources, e.g., those that arise from independent articulators, or those grounded in articulation in contrast to perception. Bias independence is modeled within the simulation by applying multiple biases that apply to one output in random order. To probe the interaction of these biases with each other and with analogical error, we will compare the stable states of three different lexicons containing distinct sets of lexical entries, where each lexicon has 50 categories corresponding to each given combination of {a, b, x}. The three lexicons are given below, where as before, an additional random category is included in each entry.

Lexicon 1: /ax/, /xb/
Lexicon 2: /axb/
Lexicon 3: /ax/, /xb/, /axb/

Figure 6 below shows a 10,000 round simulation using Lexicon (1), showing that entries for all words containing {a, x} and {b, x} maintain an average /x/ value near [+1] [13]. If we examine the relative strengths of the external biases used, we can see why this is: the tendency to shift towards +1 provided by [x] is stronger than the bias toward [-1] provided by [a] or [b], so each lexical entry has a small net bias toward [x: +1] in production.

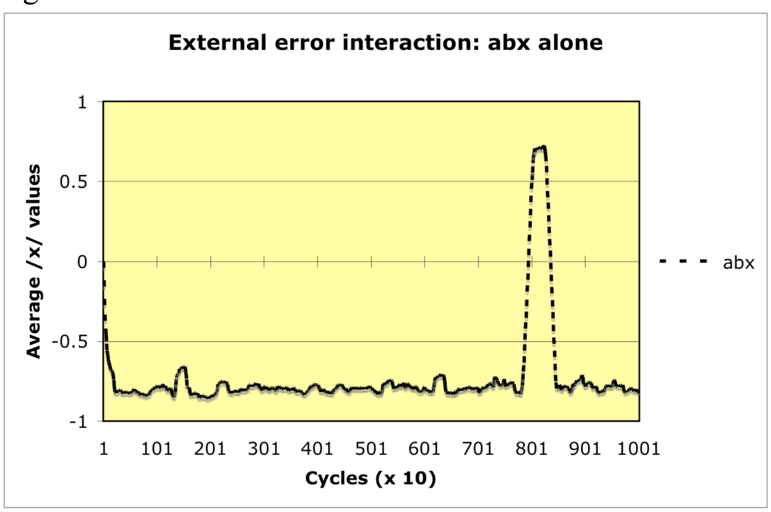
Figure 6

[13] To allow the lines in Figures 6, 7 and 8 to be seen more clearly against one another, /x/ values have been averaged over a window of 20 cycles.



However, when [a] and [b] are present together in the same output as in Lexicon 2, their independent biases should be able to interact additively and overwhelm the bias from [x]. This is shown in Figure 7, where we see that the joint influence of [a] and [b] together on output form is greater than that of [x], leading to a stable lexicon-wide /x/ value near -1.

Figure 7

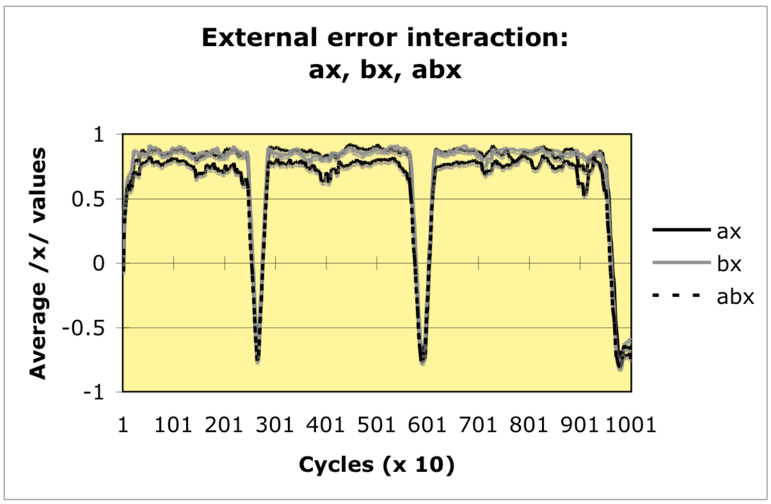


But what happens when the two patterns are put in conflict with one another in Lexicon (3) where all three types of entries are present? All lexical entries contain an /x/ category, and two pairs out of the three sets share an /a/ or /b/ category, so analogical error has a number of similarities over which to pattern-smooth. And as a consequence of this pattern smoothing, in Figure 8 we see that in the context of /a, x/ and /b, x/ entries, the biases acting on /a, b, x/ entries no longer appear additive at the level of the lexicon, even though at the level of actual error they are in fact additive[14]. Because the stronger bias on /x/ values in the /a, x/ and /b, x/ entries tends to push these outputs toward [x: +1], the /a, b, x/ outputs are pulled toward [x: +1] as well through persistent analogical error on the

[14] This additivity can be seen in the lower average /x/ value for the set of words containing both [a] and [b].

basis of shared internal structure. As a result, the entries of the form /a, b, x/ come to conform to the larger pattern set up by the forms /a, x/ and /b, x/, despite the fact that in production the net error bias for /a, b, x/ consistently pushes in the opposite direction. This pattern conforms to the OT principle of strict domination, without stipulating that the actual biases involved cannot interact additively. In this model then, a strict domination pattern is not a fundamental property of the grammar, but is an emergent result of positive feedback that tends to arise when lexical forms in which multiple constraints conflict are less frequent than forms in which subsets of those constraints conflict. This property of the model predicts that independent biases targeting distinct features of an output will often fail to pattern as additive unless there is a strong association in the lexicon between those features.

Figure 8



4.0 Modeling edge alignment in stress systems

In the previous section, we saw how positive feedback at the level of a category /x/ favored consistent outputs from that category, despite conflicting contextual biases toward production of different variants of /x/. In this section, we will make the simulation slightly more complex so that it can illustrate feedback-driven resolution of a structural conflict that isn't directly created by external bias: initial versus final edge alignment of stress. To do this, we need to modify the simulation to accommodate entries with ordered stress bearing units, i.e., syllables.

The lexicon that will be fed to the simulation contains entries consisting of ordered syllables which themselves comprise a set of unordered features. The simulation assumes an intrinsic granularity of articulatory gestures which produces a natural category boundary between ordered unit categories in speech (Goldstein et al. in press). Every syllable at the beginning of a word contains an initial-edge feature /I/, and end-syllables correspondingly contain a final-edge feature /F/. Because every word contains an initial and a final edge, they are natural targets for generalization via analogical error. The feature that is allowed to vary in its value in this simulation is stress, which is limited to the values [Stress: +1] and [Stress: -1], representing 'stressed' and 'stressless' respectively. Although edges and stress levels are not accorded the status of features in standard phonological feature systems, both of these properties condition myriad phonological patterns and so must be available at some level to the grammar/pattern-reproduction system. In a rich memory model such as this, these properties will be recorded at some level in the lexicon even if they are predictable. As before, an additional dummy feature from the set {d, e, f, ..., w} is randomly assigned to each syllable to supply additional random similarities and differences between words. For example, a possible word is /[I, d, +1] [k, -1] [F, g, +1]/, where square brackets separate syllables. As before, production and perception is modeled by retrieving the features for a given lexical entry and restoring them, where analogical or external error may intervene to change the stress value before restorage. As before, analogical error acts to bias production outputs toward the pattern of other sequences within the lexicon in proportion to similarity and frequency. As a consequence, when a particular stress pattern begins to be consistently associated with other features in the lexicon, analogical error-based feedback will reinforce this tendency until it becomes a regular pattern throughout the lexicon. The algorithm for calculating analogical error in this simulation is described in more detail in the appendix. The simulation in this section contains only one external bias, operating to promote distinct stress values in adjacent syllables at a rate of 10%. This bias ensures that stress remains alternating[15].

The lexicon that we start with contains 20 each of two, three and four syllable words, starting out with randomly assigned stress values in each syllable. Within this simulation there are four equivalently stable states that this lexicon can reach, corresponding to a particular stress value consistently aligned to the initial or the final edge. Stated in terms of feet type, this corresponds to initial- or final-aligned trochees or iambs[16]. In 20 independent runs of the simulation, all four predicted stable stress patterns emerged within 1000 cycles as expected. The stress patterns associated with the four stable states are given below in Figure 9, where stress value is abbreviated as '+' or '-'.

Figure 9. Stress alignment at cycle 1000

Stress Pattern	2 syllable words	3 syllable words	4 syllable words	Frequency out of 20
Final-aligned, Trochaic	+ –	– + –	+ – + –	4
Initial-aligned, Trochaic	+ –	+ – +	+ – + –	6
Final-aligned, Iambic	– +	+ – +	– + – +	7
Initial-aligned, Iambic	– +	– + –	– + – +	3

How does this consistent alignment to one edge happen? Analogical error results in the spread and stabilization of consistent associations between available features. However, conflict develops within the simulation because the lexicon contains both even- and odd-number syllable words, with the result that as long as stress remains alternating, there

[15] To investigate the ability of this model to reproduce patterns in which stress is not strictly alternating such as in unbounded stress systems, this bias could be split up into separate biases against clash and lapse, with different associated error rates.
[16] Addition of external error biases against particular associations would be able to bias this distribution toward that which is found crosslinguistically, in which, for example, final-aligned iambic systems appear to be rare (Hayes 1995: 262-266).

cannot be a consistent stress value associated with *both* the initial and the final edge features of all words. As we saw before however, if a given association attains a sufficient foothold in the lexicon by chance, analogical error can kick in, eventually promoting this association throughout the lexicon. In the final-aligned, trochaic system shown in Figure 9 for example, a negative stress value is strongly associated with the final edge feature, with the result that if random error happens to introduce a stressed syllable at an initial edge in one cycle, it has a good chance of being restored to negative stress in the next. On the other hand, there is no strong association of the *initial* edge feature with a particular stress value, leaving the bias toward alternating stress as the primary error influencing positions other than the final edge. As a consequence negative stress remains fixed at the final edge of each word by virtue of positive feedback, and the remaining stresses in the word fall into line from end to beginning through the bias toward alternation.

4.1 Modeling the interaction of edge-aligned stress and syllable weight

Within this model, external biases on the evolution of each lexical entry have a cumulative influence on the evolution of the entire lexicon through feedback. As a result, the frequency of category types and sequences should play a significant role in the evolution of grammatical patterns. Stated slightly differently, this model predicts that all else being equal, the larger the number of lexical entries a bias potentially applies to, the more likely its effects will come to be reflected in the behavior of the lexicon as a whole.

Gordon (1999, 2002) reports precisely such a phenomenon in his work on the relationship between the proportion of sonorants in a language's coda inventory and the likelihood that coda consonants contribute to syllable weight. Out of a sample of 400 languages that display quantity sensitive stress, Gordon found that the two most common weight distinctions are those that treat (i) CVV, but not CVC syllables as heavy, or (ii) both CVV and CVC syllables as heavy. Based on the range of phenomena that are crosslinguistically correlated with weight, as well as results of his own phonetic study, Gordon concluded that total acoustic energy in the rime is the most significant phonetic factor associated with a phonological heavy/light distinction for stress. Gordon found that the occurence of these two weight-distinctions was not random, but remarkably predictable based on the relative number of high-energy to low-energy consonants in the coda inventory. Out of a set of 24 languages in which the sonorant/obstruent and voiced/voiceless ratios in the coda inventory were less than one, 20 languages show the CVV = heavy pattern, while out of a set of 23 languages in which the sonorant/obstruent and voiced/voiceless ratios in the coda inventory were greater than one, 22 show the CVV, CVC = heavy pattern. Given that sonorants characteristically have more energy than obstruents, and voiced sounds more than voiceless sounds, this finding is consistent with the distributional data suggesting that acoustic energy in the rime is a significant factor influencing the development of quantity-sensitive stress patterns.

But this line of reasoning leaves a big question unanswered. If, for example, as suggested by the influence of sonorant to obstruent coda ratio on the probability of having CVC as heavy, sonorant codas make good stressed syllables and obstruent codas make bad ones, why don't languages develop quantity sensitive stress that targets CVV, and CV[+son] syllables, rather than just CVV and CVC? There are languages that display just this sort of split (e.g., Kwakwala, Inga Quechua (Zec 1988)) again suggesting that this reasoning is on the right track, but this pattern remains relatively rare. Gordon explains this rarity as a consequence of the pattern's greater phonological complexity and specificity relative to the more general CVV and CVC pattern. In the following section, we will see how these patterns can arise through the feedback-driven mechanism proposed here.

To model the influence of sonorant and obstruent coda consonants on stress patterns, we add to our inventory of features the set S(onorant), O(bstruent) and C(onsonant), and include them in a single syllable of a subset of lexical items in a lexicon containing 20 each of two and three syllable words, as illustrated below in Figure 10. In the starting lexicon, each word class contains equal numbers of words with initial and final-aligned stress. In addition to the external bias toward alternating stress, we include a external bias toward stressing syllables containing the feature /S/ at a rate of 5%. We do not need to specify that any of the additional features are part of a coda, because we do not overtly include vowels or onsets[17].

Figure 10: Low-frequency sonorant-coda lexicon[18]

a. Two syllable words

	Syllable 1	Syllable 2	Number of lexical entries
1	[I , Stress]	[F, Stress]	8
2	[I, O, C, Stress]	[F, Stress]	4
3	[I, Stress]	[F, O, C, Stress]	4
4	[I, S, C, Stress]	[F, Stress]	2
5	[I , Stress]	[F, S, C, Stress]	2

b. Three syllable words

	Syllable 1	Syllable 2	Syllable 3	Number of lexical entries
1	[I , Stress]	[Stress]	[F, Stress]	8
2	[I, O, C, Stress]	[Stress]	[F, Stress]	4
3	[I, Stress]	[O, C, Stress]	[F, Stress]	4
4	[I, S, C, Stress]	[Stress]	[F, Stress]	2
5	[I, Stress]	[S, C, Stress]	[F, Stress]	2

Forty percent of the words in this lexicon are just like those in the previous simulation, containing no coda of any kind (row 1 in each table). Another forty percent contain an obstruent coda (rows 2 and 3), and the remaining twenty percent contain a sonorant coda (rows 4 and 5). All words contain initial and final edge features (I and F, respectively), so

[17] In a more complex simulation which specified onsets, nuclei and codas, the external error bias could be written to only target syllables with a [VS] sequence, and analogical bias would come to build an association between stress and consonant features that have a preceding V.

[18] The value of the stress feature, /Stress: +1/ or /Stress: -1/, is random in the initial cycle, so no value is marked in this depiction of the initial lexicon.

just as before, feedback error will promote development of a consistent alignment of a stress value to an edge. Since the sonorant codas are split across the lexicon between syllable 1 and 2, for ten percent of the lexicon a constant error bias towards stressing syllables with sonorant codas will act to disrupt any edge aligned pattern. While the edge aligned pattern can be stabilized by associations that are present in every word, the only opportunity for a sonorant-coda stressing pattern to be stabilized by feedback is if a consistent association between the feature /S/ and /Stress: +1/ can develop in the twenty percent of words that contain this feature. This association would also create an incipient association between /C/ and /Stress: +1/ within the words containing a sonorant coda, but because these words are in the minority of all words containing /C/, positive feedback is unlikely to gain a foothold to induce the larger number of words containing an obstruent coda to follow suit. And out of twenty independent runs of this simulation, at 1000 cycles eighteen showed one of the four possible lexicon-wide consistent edge alignments, and two were caught between patterns, indicating that a positive feedback driven association of stress to an edge in all words can overcome a strong external bias toward stressing sonorant codas in a few words.

Given the role of consistency of patterning in promoting positive feedback in this model, this result is not surprising. The constant bias to stress syllables with sonorant codas applies to few words in the lexicon, and these few words share many features with other words that are not affected by this constant bias, meaning that words containing sonorant codas will always be pulled away toward the majority pattern by positive feedback over other associations. Naturally what we should try next is a lexicon with a greater proportion of words with sonorant codas relative to those with obstruent codas, as shown in Figure 11. This lexicon is identical to that shown in Figure 10, except that the number of lexical entries in rows 2 and 3 have been halved, and that in rows 4 and 5 have been doubled. As a result, now twice as many words have sonorant codas as obstruent codas, instead of vice versa.

Figure 11: High-frequency sonorant-coda lexicon

a. Two syllable words

	Syllable 1	Syllable 2	Number of lexical entries
1	[I , Stress]	[F, Stress]	8
2	[I, O, C, Stress]	[F, Stress]	2
3	[I, Stress]	[F, O, C, Stress]	2
4	[I, S, C, Stress]	[F, Stress]	4
5	[I , Stress]	[F, S, C, Stress]	4

b. Three syllable words

	Syllable 1	Syllable 2	Syllable 3	Number of lexical entries
1	[I , Stress]	[Stress]	[F, Stress]	8
2	[I, O, C, Stress]	[Stress]	[F, Stress]	2
3	[I, Stress]	[O, C, Stress]	[F, Stress]	2
4	[I, S, C, Stress]	[Stress]	[F, Stress]	4
5	[I, Stress]	[S, C, Stress]	[F, Stress]	4

When twenty independent runs of this simulation were run out to 1000 cycles, three different patterns emerged, plus a residue of inconsistently patterned lexicons in transition between states. I will remark on each of these patterns in turn.

Four out of the twenty lexicons showed a pattern of stressing all syllables with an /S/ feature, while the syllables without an /S/ feature showed a consistent alignment of a stress value to an edge. Clearly, increasing the number of syllables with sonorant codas in the lexicon to 40%, in conjunction with a strong constant bias toward stressing these syllables, allowed the development of a sufficiently strong association between /S/ and /+/ to maintain a separate pattern from the rest of the lexicon, which itself went ahead and aligned consistently to an edge. These outcomes parallel those languages that stress CV: and CV[+son] but not CV[-son]. Interestingly however, this is not the majority pattern, for another eleven out of the twenty lexicons showed a pattern of stressing *all* syllables with a coda, without regard to whether that coda was sonorant or obstruent, with the remaining 40% of the words consistently showing one of the four edge alignments. This occurs for two reasons. Recall that of the syllables containing /C/ in the high-frequency sonorant lexicon, 67% also contain /S/, and that at any given time about half of the remaining syllables containing /C/ will also be stressed because stress is alternating. Consequently, as soon as all sonorant codas are stressed, more than 80% of all syllables containing /C/ are stressed, creating a strong basis for positive feedback to promote the development of stress in all syllables containing /C/. Secondly, as soon as all syllables containing sonorant codas become stressed, the edge alignment pattern becomes weakened, because half of the words containing sonorant codas will violate any established edge alignment. The combination of weaker positive feedback maintaining alignment to an edge, combined with the strong incidental association of /C/ with /Stress: +1/ allows development of a lexicon in which all syllables with a coda are stressed, even though the actual bias promoting this pattern only applies to a subset of those syllables. Finally, two of the twenty lexicons showed lexicon-wide edge-alignment of stress, and three did not show a clear pattern.

The wide range of patterns arising in this simulation illustrates several issues in a model of a rich lexicon evolving through positive feedback. First, positive feedback that generalizes beyond the source pattern must make use of categories, whether present innately or abstracted from the input data (Maye et al. 2002, Saffran and Thiessen 2003, reviewed in Gerken 2006). In this case, the fact that the simulation can generalize to codas from sonorant codas requires, as asserted by many featural systems, that a separate category of ‘consonant’ be available which can serve as the basis for generalizing positive feedback. Second, the type frequency of categories and category associations in the lexicon matters. Positive feedback is definitionally auto-catalytic, such that the ability of a positive feedback mechanism to influence a process is constrained by the ability of an event to catalyze a subsequent event. If an external bias only affects a small proportion of a set of lexical items linked by similarity in a number of features, noise in production/perception operating over the entire set of lexical items will lower the probability that this local bias can initiate a set-wide change. As clear from the simulation above, and in Gordon’s empirical findings on the influence of the coda-inventory on weight systems, relative frequency of the target for a bias within a similarity set has a significant role in influencing the effect of the bias. As a consequence, this model predicts that a strong bias against a low-frequency combination of features may be less likely to initiate an active phonological pattern than a weaker bias against a higher-frequency combination.

Once entrenched, patterns in the lexicon can be self-sustaining through continuous positive feedback. Given that positive feedback also promotes coarse- over fine-grained patterns, this provides a potential explanation for observed divergent allophonic context-dependencies. As an illustration, let us return briefly to the example of context-dependent lateral allophones. Let us assume, for the sake of argument, that there are external biases favoring plain (versus velarized) laterals in the syllable onset, and that there are independent external biases favoring plain laterals before front vowels. As illustrated above, positive feedback is likely to favor the development of a simple contextual condition for plain allophones of /l/. This means that all else being equal a plain allophone is more likely to either appear in onset position (as in English) or preceding a front vowel (as in Georgian), but less likely to appear in a context defined by an interaction of the two independent conditions. Standard Optimality Theory accounts for divergent patterns in different languages by positing the existence of, and language-specific ranking of innate context-specific constraints. In contrast, the model presented here accounts for such divergent patterns through the operation of positive feedback over external biases targeting distinct contexts.

5.0 Conclusions

The model described in this paper is intended to contribute to our understanding of how regularity might arise within a rich-memory lexical system without recourse to highly specified, innate restrictions. The model rests on the self-organizing interaction of two conceptually distinct sources for error in production and perception to produce regular sound patterns. Analogical error, operating at potentially many levels, results in a tendency to produce and perceive forms with a bias towards previous experienced forms in some relation to similarity and frequency. Because errors of this type increase similarity over time, they push the system toward categorical behavior even in the absence of any intrinsic or extrinsic bias toward that behavior.

A second conceptual source of error can be idealized as relatively independent of experience, deriving from more constant facts-about-the-world such as physiological constraints on sound production or perception. In simulations, we saw that external error produces statistical tendencies within the lexicon that can initiate runaway positive feedback from analogical error, resulting in the consistent development of particular kinds of regular patterns across the lexicon. As a consequence, the lexicon tends to evolve under the influence of analogical error to categorically recapitulate the gradient external biases that successfully initiate asymmetries in the first place. However, once a categorical pattern has developed, maintenance of the pattern may become largely independent of external error, and may be extended to similar forms that are not targets of the error that originated the pattern (see, e.g., Mielke 2004, pp 92-140 for discussion and review of this point). Furthermore, this model predicts that any asymmetry in experience, whether prompted by 'natural' phonetic factors or not, can serve as a nucleation point for feedback-driven development of a regular pattern. As a result, we expect many languages to share regular patterns that derive from common causes in physiology, but we also expect variously idiosyncratic patterns to arise in response to historically contingent factors. This supports Blevins' conceptual division of patterns into 'natural' and 'unnatural' classes, where the first derive from lower-level, relatively constant features of human speech production and processing modeled by the external biases in the simulations presented here, and the second from more unusual confluences of events (Blevins 2005, 2006).

It has been frequently noted that phonological patterns appear to be coarser in a number of ways than potentially possible given the more gradient phenomena that underlie them (e.g., Pierrehumbert 2001b, 2003, Gordon 2002, discussed in section 3.3 above). Pierrehumbert (2001b, 2003) offers the argument that the relative 'coarseness' of phonological patterns results from learnability constraints in the presence of noise: possible phonological constraints that are highly specific in terms of the phonetic cues they refer to will necessarily refer to a smaller set of the material over which induction occurs, and therefore are more susceptible to interference from noise in the signal.

The model used here presents complementary, conceptually similar mechanisms that function to encourage consistency and generality in grammatical patterns. First, external biases that refer to highly specific contexts will target relatively fewer forms, and therefore will be less likely than more general biases to initiate a consistent, productive pattern in the lexicon – even though these highly specific biases may be detectable at a phonetic level. Second, if a pattern does manage to take hold over a small subset within a larger similarity set, it is likely to either extend its pattern to the larger set, thereby becoming more general (e.g., the extension of stress to all coda consonants in section 4.1 above), or to be reabsorbed into the pattern of the larger set (as in Figure 4 above). This constant drift toward pattern consolidation through analogical error mitigates the tendency of highly context sensitive biases in performance to produce ever greater distinctions in lexical form within a rich-memory model of the lexicon. The conflict between the two tendencies should result in a shifting compromise in which the phonology often exhibits fewer and more regular distinctions than are potentially motivated by the phonetics.

This paper focuses on positive feedback as a source of regularity in the lexicon. Considerable evidence suggests that mechanisms of language production and perception create systematic error biases toward increasing similarity between forms. Under the influence of these biases, small asymmetries in the distribution of lexical properties become attractors in subsequent cycles of production and perception, and over longer time scales, in transmission. The initial asymmetries that are exaggerated by feedback may be random, or conditioned by more constant, system external biases, such as physiological constraints on voicing. The central finding is that under feedback, error biases toward similarity results in emergent regularity of sound patterns across the lexicon. Though the results presented involve simulations, there is growing empirical evidence in the phonetics/phonology literature for feedback systems of this kind. Within the tradition of laboratory phonology, Ohala (1981, 1989, 1990) has stressed the parallelism between common misperception and 'mini-sound changes'. Misperception in the laboratory demonstrates system-external biases which parallel crosslinguistically common regular sound patterns. In the model presented here, these persistent biases in misperception can provide the initial nucleation point for the development of categorical pattern.

Within the tradition of historical and theoretical phonology, there has also been a recent renewal of interest of the role of system-internal attractors in sound change. Though relationships between pre-existing sound patterns and sound change were noted in early work on compensatory lengthening (de Chene and Anderson 1979), it is only recently that attempts have been made to integrate these into a more comprehensive theory of sound change (Blevins 2004:153-55, 247-48, 297-99; Blevins to appear b). The results strongly support this program. For example as demonstrated by Chitoran and Hualde (this volume), there are clear correlations between the shift of hiatus to glide-vowel sequences and the pre-existence of glide-vowel sequences in the lexicons of different Romance languages. System-internal attractors are also clearly in evidence in Austronesian, as argued by Blust (this volume), where disyllabic word canons appear to act as attractors for a range of independent sound change types, including vowel loss between identical consonants. Under this model, synchrony describes the current state of the system. However, given clear evidence that even an adult system is continually modified by experience, the opportunity for feedback renders each current state 'diachronic' as soon as new input is encountered. A synchronic grammar must therefore be considered an abstraction, rather than a description of a true steady state. Instead, a grammar appears to be a complex evolving system of competing regularities and generalizations in which change may be rapid or slow depending on the details of the current system and environment. Modeling sound systems in this way will likely bring us closer to understanding the true nature of sound patterns, and their evolution over time.

Appendix: Modeling analogical error algorithmically

The algorithm used here is not based on any concrete model of how such an analogical error bias might actually function in language production or perception. Rather, it represents a simple and computationally tractable way to identify common associations between features in the current lexicon, and then accordingly weight the probability of error in the production of a particular lexical item sharing those features. In each cycle, the entire set of associations between all category labels present in all lexical items and the value of /x/ is identified. To do this, within the simulations in section 3 the power set of the category labels present in each lexical entry is calculated and each subset (except for the empty set) is associated with the value of /x/ in that lexical entry. For example, for a lexical entry /a, b, x: +1/, the existing associations are:

/a/ ↔ +1
/a, b/ ↔ +1
/a, b, x/ ↔ +1
/a, x/ ↔ +1
/b/ ↔ +1
/b, x/ ↔ +1
/x/ ↔ +1

The /x/ values of all associations over the entire lexicon that share a left-hand side are added together, resulting in an expression that provides a measure of the consistency of association with a given category label set with a particular value of /x/. For example, if within the lexicon, the label /a/ is strongly associated with [x: +1], the summed associations with the feature sets /a/, /a, b/, /a, b, x/ and /a, x/ will be strongly positive. On the other hand, if the feature /b/ is not associated with any particular value of /x/, there will be about as many associations of /b/ with [x: +1] as [x: -1], and the summed associations that contain /b/ but not /a/ will cluster around zero. See Broe 1996, Albright 2002, Wedel 2004 and Bod 2006 for additional examples of the use of the power set of a group of linguistic elements to discover consistent associations between any subsets of the group.

When a lexical entry is produced, analogical error is allowed to influence to the output 10% of the time; otherwise, the output is produced with external and random error alone. When analogical error contributes, the summed associations with /x/ for all combinations of labels in the lexical entry under production are themselves summed together. This final sum is a measure over the lexicon of the overall bias of the sequences contained in the lexical entry toward a positive or negative value for /x/. If the resulting value is greater than zero, [x] is set at [+1]; if it is less than zero, [x] is set at [-1]. For strong associations across the lexicon, the total sum representing the lexical internal bias for particular lexical entries will be consistently positive or negative, while if associations are weak, they will tend to oscillate back and forth from positive to negative, resulting in little net change over time.

For the more complex simulations involving ordered category sets in section 4, the algorithm was the same, except that associations were not calculated over an entire entry, but only over the syllable of the /stress/ value under production and the syllable immediately preceding and following. This limitation was imposed for computational tractability.

References

Albright, Adam (2002). The lexical bases of morphological well-formedness. In Sabrina Bendjaballah, Wolfgang Dressler, Oskar Pfeiffer and Maria Voeikova, (eds.) *Morphology 2000*. Philadelphia: Benjamins.

Albright, Adam and Bruce Hayes (2002). Modeling English Past Tense Intuitions with Minimal Generalization. In Michael Maxwell, (ed.) *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology*. Philadelphia, July 2002. ACL.

Alegre, Maria and Peter Gordon (1999) Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41-61.

Almor, Amit (in press). Past tense learning. In Michael Arbib (ed.) *The Handbook of Brain Theory and Neural Networks*, Second edition, Cambridge, MA: The MIT Press.

Anderson, Stephen. R. (1985). *Phonology in the Twentieth Century: Theories of Rules and Theories of Representations*. Chicago and London: The University of Chicago Press.

Archangeli, Diana and Douglas Pulleyblank (1994). *Grounded Phonology*. Cambridge: MIT Press.

Barton, David (1980). Phonemic Perception in Children. In Grace. H. Yeni-Komshian, James. F. Kavanagh, and Charles. A. Ferguson, (eds.) *Child Phonology: Volume 2, Perception*. New York: Academic Press.

Batali, John (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In Ted Briscoe, (ed.) *Linguistic evolution through acquisition*. Cambridge: Cambridge University Press.

Blevins, Juliette (2004). *Evolutionary Phonology: The emergence of sound patterns.* Cambridge: Cambridge University Press.

Blevins, Juliette (2005) Understanding antigemination. In Zygmunt Frajzyngier (ed.), *Linguistic Diversity and Language Theories*. Amsterdam/Philadelphia, John Benjamins, 203-234.

Blevins, Juliette. (2006). New perspectives on English sound patterns: 'Natural' and 'Unnatural' in Evolutionary Phonology. *Journal of English Linguistics* 34:6-25.

Blevins, Juliette (To appear a). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*.

Blevins, Juliette (To appear b). Phonetically-based sound patterns: Typological tendencies or phonological universals? *Proceedings of LabPhon 10*. Mouton de Gruyter.

Blevins, Juliette, and Robert Blust (2003). Towards a theory of drift: Patterns of Austronesian syncope. Ms. University of California, Berkeley.

Blevins, Juliette and Andrew Garrett (1998). The origins of consonant-vowel metathesis. *Language* 74: 508-56.

Bod, Rens (2006). An All-Subtrees Approach to Unsupervised Parsing. *Proceedings of ACL-COLING*, Sydney, Australia.

Brighton, Henry, Simon Kirby, and Kenny Smith (2005). Cultural Selection for Learnability: Three Principles Underlying the View that Language Adapts to be Learnable. In Maggie Tallerman, (ed.) *Language Origins: Perspectives on Evolution*. Oxford: Oxford University Press.

Broe, Michael (1996). A generalized information-theoretic measure for systems of phonological classification and recognition. *Computational Phonology in Speech Technology: Second meeting of the ACL Special Interest Group in Computational Phonology, Proceedings of the Workshop*, pp. 17-24.
Burzio, Luigi (2005). Sources of Paradigm Uniformity. In Laura J. Downing, T. Alan Hall, Renate Raffelsiefen, (eds.), *Paradigms in Phonological Theory*, pp. 65-106. Oxford: Oxford University Press.
Bybee, Joan (1985). *Morphology: A study of the relation between meaning and form*. Amsterdam: J. Benjamins.
Bybee, Joan (2001). *Phonology and language use*. Cambridge: Cambridge University Press.
Bybee, Joan (2002a). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14: 261-290.
Bybee, Joan (2002b). Phonological Evidence for Exemplar Storage of Multiword Sequences. *Studies in Second Language Acquisition* 24: 215-221.
Camazine, Scott, Jean-Louis Deneubourg, Nigel Franks, James Sneyd, Guy Theraulaz, and Eric Bonabeau (2001). *Self-Organization in Biological Systems*. Princeton: Princeton University Press.
Carter, Paul G. (1999). Abstractness in phonology and extrinsic phonetic interpretation: the case of liquids in English. *Proceedings of the 14 International Congress of Phonetic Sciences*: 105-108.
Chomsky, Noam and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper and Row.
Cziko, Gary (1995). *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*. Cambridge, MA: MIT Press.
de Boer, Bart (2001). *The origins of vowel systems*. Oxford: Oxford University Press.
De Chene, Brent and Stephen R. Anderson. 1979. Compensatory lengthening. *Language* 55: 505-35.
Dell, Gary S., Kristopher D. Reed, David R. Adams, and Antje S. Meyer (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26: 1355–1367.
Donegan, Patricia and David Stampe (1979). The Study of Natural Phonology. In D. A. Dinnsen (ed.), *Current Approaches to Phonological Theory*. Bloomington: Indiana University Press.
Ernestus, Mirjam and R. Harald Baayen (2003). Predicting the unpredictable: Interpreting neutralized segments in Dutch. *Language* 79, 5-38.
Frisch, Stephen (1996). *Similarity and Frequency in Phonology*. Ph.D. Dissertation, Northwestern University.
Gerken, LouAnn A. (2006). Decisions, decisions: infant language learning when multiple generalizations are possible. *Cognition* 98: B67-B74.
Goldinger, Stephen D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning Memory and Cognition* 22: 1166-1182.

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

Goldstein, Louis, Dani Byrd, and Elliot Saltzman (in press). The role of vocal tract gestural action units in understanding the evolution of phonology. In Michael Arbib, (ed.) *From Action to Language: The Mirror Neuron System*. Cambridge: Cambridge University Press.

Gordon, Matthew (1998). The phonetics and phonology of non-modal vowels: a cross-linguistic perspective. *Proceedings of the Berkeley Linguistics Society* 24, 93-105

Gordon, Matthew (1999). *Syllable weight: phonetics, phonology, and typology*. Los Angeles: UCLA Ph.D. dissertation.

Gordon, Matthew (2002). A phonetically-driven account of syllable weight. *Language* 78: 51-80.

Guenther, Frank H., and Gjaja, Marin. N. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America* 100, 1111-1121.

Guenther, Frank H., Alfonso Nieto-Castanon, Satrajit S. Ghosh, and Jason A. Tourville (2004). Representation of sound categories in auditory cortical maps. *Journal of Speech, Language, and Hearing Research* 47, 46-57.

Hare, Mary. and Jeffrey L. Elman (1995). Learning and morphological change. *Cognition* 56, 61-98.

Harrington, Jonathan, Sallyanne Palethorpe and Catherine I. Watson. (2000). Does the Queen speak the Queen's English? *Nature* 408, 927-928.

Hayes, Bruce (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: University of Chicago Press.

Hayes, Bruce (1999). Phonetically-Driven Phonology: The Role of Optimality Theory and Inductive Grounding. In Michael Darnell, Edith Moravscik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly, (eds.), *Functionalism and Formalism in Linguistics, Volume I: General Papers*. pp. 243-285. Amsterdam: John Benjamins.

Hayes, Bruce and Donca Steriade (2004). The phonetic bases of phonological markedness. In Bruce Hayes, Robert Kirchner, and Donca Steriade, (eds.). *Phonetically Based Phonology*. pp. 1-33. Cambridge: Cambridge University Press.

Hickey, Raymond (2004). Irish English Phonology. In Edgar Schneider, Kate Burridge, Bernd Kortman, Mesthrie Rajend and Clive Upton, (eds.) *A Handbook of the Varieties of English,* Vol. 1, pp 70-97. New York: Mouton de Gruyter.

Hintzman, Douglas (1986). "Schema Abstraction" in a Multiple-Trace Memory Model. *Psychological Review* 93: 411-428.

Hock, Hans H. (1991). *Principles of Historical Linguistics* (2nd ed.). The Hague: Mouton de Gruyter.

Hyman, Larry (1977). Phonologization. In A. Juilland (ed.), *Linguistic studies presented to Joseph H. Greenberg*, pp. 407-418. Saratoga: Anna Libri.

Ito, Junko and Armin Mester (2003). Japanese Morphophonemics: Markedness and word structure. Cambridge, MA: MIT Press.

Jacoby, Larry. L. and Lee Brooks (1984). Nonanalytic cognition: Memory, perception and concept learning. In Gordon H. Bower, (ed.), *The psychology of learning and motivation: Advances in research and Theory*, Vol. 18. San Diego, CA: Academic Press. 1-47.

Jakobson, Roman, Gunnar Fant, and Morris Halle (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.
Joanisse, Marc F. and Mark 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, Hilsdale, NJ: Erlbaum. 331-336.
Johnson, Keith (1997). Speech perception without speaker normalization. In Keith Johnson and John W. Mullennix (eds.) *Talker Variability in Speech Processing*. San Diego: Academic Press. 145-166.
Johnston, Paul (2002). Regional Variation. In: Charles Jones (ed.) *The English Language in Scotland: An introduction to Scots*. East Clinton: Tuckwell.
Joseph, Brian and Richard Janda (1988) The How and Why of Diachronic Morphologization and Demorphologization. In Michael Hammond and Michael (eds.), *Theoretical Morphology: Approaches in Modern Linguistics*. San Diego CA: Academic Press.193-210.
Joseph, Brian and Richard Janda (1997) How General are our Generalizations? What Speakers Actually Know and What They Actually Do. In Anthony D. Green and Virginia Motopanyane, (eds.), ESCOL 96. *Proceedings of the Thirteenth Eastern States Conference on Linguistics*. Ithaca: Cascadilla Press. 148-160.
Kaufmann, Stuart A. (1995). *At home in the universe: The search for the laws of self-organization and complexity*. Oxford: Oxford University Press.
Kaun, Abigail (1995) *The Typology of Rounding Harmony: An Optimality Theoretic Approach*. Los Angeles: UCLA Ph.D dissertation.
Kelso, J. A. Scott. (1995) *Dynamic patterns: The self-organization of brain and behavior.* Cambridge, MA: MIT Press.
Kenstowicz, Michael (1994). *Phonology in Generative Grammar*. Cambridge, MA and Oxford: Blackwell.
Kirby, Simon and James Hurford (2002). The Emergence of Linguistic Structure, An Overview of the Iterated Learning Model. In Angelo Cangelosi and Dominic Parisi (eds.), *Simulating the Evolution of Language*. London: Springer. 121-148.
Kruschke, John. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99, 22-44.
Kuehne, Sven. E., Kenneth D. Forbus, Dedre Gentner and Brynn Quinn (2000). SEQL - Category learning as progressive abstraction using structure mapping. *Proceedings of the Twenty-second Annual Conference of the Cognitive Science Society*, 770-775.
Kuhl, Patricia 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.
Kuhl, Patricia. K. (1995). Mechanisms of developmental change in speech and Language. In Kjell Elenius and Peter Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 2. Stockholm: Stockholm University, 132–139.
Lacerda, Franciso (1995). The perceptual-magnet effect, An emergent consequence of exemplar-based phonetic memory. In Kjell Elenius and Peter Branderud (eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Vol. 2. Stockholm: Stockholm University, 140–147.

Labov, William (1994). *Principles of linguistic change, Vol. 1, Internal Factors*. Oxford and Cambridge, MA: Blackwell.
Ladefoged, Peter (2006). *A course in phonetics, 5th edition*. Boston: Thomson.
Lee, S., Potamianos, A. & Narayan, S. (1999) Acoustics of children's speech: developmental changes of temporal and spectral parameters. *Journal of the Acoustical Society of America* 105: 1455-1468.
Levelt, Willem. J. M. (1989). *Speaking*. Cambridge MA: MIT Press.
Lindblom, Björn (1998). Systemic constraints and adaptive change in the formation of sound structure. In James Hurford, Michael Studdert-Kennedy and Christopher Knight, (eds.) *Approaches to the evolution of Language*. Cambridge: Cambridge University Press.
Lindblom, Björn, Peter MacNeilage and Michael Studdert-Kennedy (1984). Self-organizing processes and the explanation of language universals. In Brian Butterworth, Bernard Comrie and Östen Dahl, (eds.), *Explanations for language universals*. Berlin: Walter de Gruyter & Co. 181-203.
Long, Christopher and Amit Almor (2000). Irregularization: The interaction of item frequency and phonological interference in regular past tense production. *Proceedings of the twenty-second annual conference of the cognitive science society*. Hillsdale, NJ: Lawrence Erlbaum Associates.
Luce, Paul. A. and David B. Pisoni (1998). Recognizing Spoken Words: the neighborhood activation model. *Ear and Hearing* 19:1-36.
MacWhinney, Brian (1998). Models of the emergence of language. *Annual Review of Psychology*, 49, 199-227.
Maye, Jessica, Janet F. Werker, and LouAnn A. Gerken (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82: B101–B111.
Mendoza-Denton, Norma (to appear). *Homegirls: Symbolic Styles in the Making of Latina Girl Gangs*. London: Blackwell.
Mielke, Jeff (2004). The Emergence of Distinctive Features. PhD Dissertation, Ohio State University.
Nosofsky, Robert (1986). Attention, Similarity, and the Identification-Categorization Relationship. *Journal of Experimental Psychology, General* 115, 39-57.
Nosofsky, Robert (1988). Similarity, Frequency, and Category Representations. *Journal of Experimental Psychology, Learning, Memory and Cognition* 14, 54-65.
Ohala, John (1981). The Listener as a Source of Sound Change. In *Proceedings of the Chicago Linguistics Society* 17, *Papers from the Parasession on Language and Behavior*, 178-203.
Ohala, John (1983). The origin of sound patterns in vocal tract constraints. In: Peter. F. MacNeilage (ed.), *The production of speech,* pp. 189 - 216. New York: Springer-Verlag.
Ohala, John (1989). Sound change is drawn from a pool of synchronic variation. In Leiv E. Breivik and Ernst H. Jahr, (eds.) *Language Change: Contributions to the study of its causes*. [Series: *Trends in Linguistics, Studies and Monographs* No. 43]. Berlin: Mouton de Gruyter.

Ohala, John (1990). The phonetics and phonology of aspects of assimilation. . In John Kingston and Mary Beckman (eds.), Papers in Laboratory Phonology I. Cambridge University Press, Cambridge. 258-75.
Ohala, John (2005). Phonetic explanations for sound patterns. Implications for grammars of competence. In W. J. Hardcastle & J. M. Beck (eds.) *A figure of speech. A festschrift for John Laver*. pp. 23-38. London: Erlbaum.
Oliphant Michael (2002). Learned systems of arbitrary reference: The foundation of human linguistic uniqueness. In Ted Briscoe (ed.) *Linguistic evolution through acquisition*. Cambridge: Cambridge University Press.
Oudeyer, Pierre-Ives (2002). Phonemic coding might be a result of sensory-motor coupling dynamics. *Proceedings of the 7th International Conference on the Simulation of Adaptive Behavior*. In Bridget Hallam, Dario Floreano, John Hallam, Gillian Hayes, Jean-Arcady Meyer, (eds.) pp. 406-416. Cambridge: MIT Press.
Padgett, Jaye (2002). Constraint conjunction versus grounded constraint hierarchies in Optimality Theory. Ms. University of California, Santa Cruz.
Pierrehumbert, Janet (2001a). Exemplar dynamics: Word frequency, lenition, and contrast. In Joan Bybee and Paul Hopper (eds.) *Frequency effects and the emergence of linguistic structure*. John Benjamins, Amsterdam. 137-157.
Pierrehumbert, Janet (2001b). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes* 16: 691-698.
Pierrehumbert, Janet (2002). Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner, (eds.) *Laboratory Phonology* 7. Berlin; New York : Mouton de Gruyter.
Pierrehumbert, Janet (2003). Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech* 46: 115-154.
Plaut, David C. and Christopher T. Kello (1999) The Emergence of Phonology From the Interplay of Speech Comprehension and Production: A Distributed Connectionist Approach. In Brian MacWhinney (ed.), *Emergence of Language*. Hillsdale, NJ: Lawrence Earlbaum Associates.
Pouplier, Marianne, Larissa Chen, Louis Goldstein and Dani Byrd (1999). Kinematic evidence of the existence of gradient speech errors. *Journal of the Acoustical Society of America*, 106, 2242.
Pouplier, Mariane and William Hardcastle (2005). A Re-Evaluation of the Nature of Speech Errors in Normal and Disordered Speakers. *Phonetica* 62, 227-243.
Prince, Alan and Paul Smolensky (1993). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report CU-CS-696-93, Department of Cognitive Science, University of Colorado at Boulder, and Technical Report TR-2, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ.
Redford, Melissa A., Chun C. Chen, and Risto Miikkulainen (2001). Constrained Emergence of Universals and Variation in Syllable Systems. *Language and Speech* 44: 27-56.
Saffran, Jenny R., and Erik D. Thiessen (2003). Pattern induction by infant language learners. *Developmental Psychology* 39: 484–494.
Sancier, Michele L. and Carol A. Fowler (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics 25*: 421-436.

Saltzman, Elliot and Kevin G. Munhall (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1: 333-382.
Shattuck-Hufnagel, Stephanie, and Dennis 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.
Skousen, Royal. (1989). *Analogical modeling of language*. Dordrecht, Boston and London: Kluwer.
Smith, Andrew D. M. (2005) The Inferential Transmission of Language. *Adaptive Behavior* 13, 311-324.
Sproat, Richard and Osamu Fujimura (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21: 291-311.
Steels, Luc (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103:1-24.
Steels, Luc and Frederic Kaplan (2002). Bootstrapping grounded word semantics. In Ted Briscoe (ed.) *Linguistic evolution through acquisition*. Cambridge: Cambridge University Press.
Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. *UCLA Working Papers in Linguistics* 3: 25-146.
Steriade, Donca (1999). Alternatives to syllable-based accounts of consonantal phonotactics. Osamu Fujimura, Brian D. Joseph, and Bohumil Palek (eds.) *Proceedings of LP '98: Item Order in Language and Speech*. pp. 205-245. Prague: Karolinum Press.
Tenpenny, Patricia. L. 1995. Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review* 2: 339-363.
Tuller, Betty, Pamela Case, Mingzhou Ding, and Scott Kelso (1994). The non-linear dynamics of categorical perception. *Journal of Experimental Psychology: Human perception and Performance*, 20: 3-16.
Trubetzkoy, Nikolai (1939). *Grundzüge der Phonologie. Travaux du cercle linguistique de Prague* 7.
Tschenkeli, Kita (1958). *Einführung in die Georgische Sprache, Band 1*, p. LIV. Zürich: Amirani Verlag.
Wedel, Andrew (2004). *Self-organization and categorical behavior in phonology*. Santa Cruz CA: University of California Santa Cruz Ph.D Dissertation.
Wedel, Andrew (in press). Exemplar models, evolution and language change. *The Linguistic Review*.
Wright, Richard and Stephan Frisch (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30, 139–162.
Zanone, Pier. G., and Kelso, J. A. Scott (1992). The evolution of behavioral attractors with learning, nonequilibrium phase transitions. *Journal of Experimental Psychology, Human Perception and Performance* 18, 403-421.
Zanone, P. G., and Kelso, J. A. Scott (1997). The coordination dynamics of learning and transfer, Collective and component levels. *Journal of Experimental Psychology, Human Perception and Performance* 23, 1454–1480.
Zec, Draga (1988). Sonority constraints on prosodic structure. PhD dissertation, Stanford University [Published by Garland 1994].