UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SELF-ORGANIZATION AND CATEGORICAL BEHAVIOR IN PHONOLOGY

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

LINGUISTICS

by

Andrew B. Wedel

March 2004

The Dissertation of Andrew B. Wedel
is approved:

____
Professor Jaye Padgett, Chair

____
Professor Armin Mester

____
Professor Adam Albright

____
Robert C. Miller
Vice Chancellor for Research and
Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

Self-organization and categorical behavior in phonology

Andrew Benjamin Wedel

Generative models of phonology account for output patterns through a complex grammar applied over a minimal lexicon. In contrast, many natural complex patterns result from the gradual accumulation of structure through repeated local interactions. In this dissertation I present results of simulations supporting the proposal that some phonological patterns can be accounted for through self-organization within an analogically structured lexicon, in response to forcing from external biases. In Chapter 1, I show that patterns accounted for by the Optimality-Theoretic principles of constraint dominance and strict constraint dominance can be shown to spontaneously arise in analogically-structured systems, driven by competition between leveling pressures within the lexicon and differentiating pressures from lexicon-external performance biases.

Phonological systems exhibit 'constrained contrast' in two distinct ways: first, phonologies exhibit only a subset of cross-linguistically attested contrasts, formed from a subset of possible features in combination. Second,

crosslinguistically infrequent elements also tend to occur less frequently in a language that does have them. In Chapter 2, I present evidence that both of these patterns can be accounted for diachronically through indirect selection over phonetic variants, given the assumptions that, 1) lexical categories are richly specified, 2) a perceived utterance updates the content of a lexical category only if it is identified as an example of that lexical category, and 3) lexical categories can influence each others' production in proportion to phonological similarity.

When a simulated speaker/hearer pair alternately communicate their lexicons to each other under these conditions their lexicons converge. Further, when an output is too close to multiple categories, it is less likely be consistently categorized, with the result that it has less influence on the evolution of the pairs' lexicons, resulting in pressure on lexical categories to remain contrastive. W hen biases against certain features or feature combinations are introduced, the pairs' lexicons evolve to avoid as many of these 'less-fit' elements as possible. However, when avoidance of all marked elements would result in insufficient contrast, the lexicons evolve to utilize a subset of less-fit elements, but at a lower frequency than fitter elements.

Acknowledgements

My dissertation advisor Jaye Padgett has played a crucial role in the inception, development and conclusion of this dissertation, and I am grateful beyond words for his support, his criticisms, and his friendship. I also owe a debt of gratitude to the other members of my committee, Armin Mester and Adam Albright, whose questions and comments throughout the process have helped me become a considerably better linguist. I am also deeply indebted to Juliette Blevins for her continuing encouragement and thoughtful criticism as I pursue the ideas explored in this dissertation. Among my fellow Santa Cruz graduate students, I particularly wish to thank Matt Chisholm, Patrick Davidson, and Afton Lewis for their great intelligence and moral support at the beginning of this project, and Brian Ort for his knowledge of evolutionary theory at the end. Nearer to my heart, I wish to thank my beloved parents and brother, who never reproached me for giving up an already launched career in molecular biology for an uncertain future in linguistics. Finally, I owe my present wonderful life to my dear partner in life and crime Adam Ussishkin, without whom I could never have done any of this at all.

Chapter 1. Self-organization and the outcome of pattern conflict

1.0 Introduction

Sound patterns in utterances vary widely in their consistency and generality. Some sound patterns are highly gradient (e.g., Keating 1988, Cohn 1990, Liberman and Pierrehumbert 1994, Smith 1997), while others are more nearly categorical (see Hayes 1999b for general discussion). For example, in languages allowing surface voiced obstruents in word-final position, voicing is often variably reduced depending on a variety of contextual influences such as phrasal position, stress and the characteristics of a following segment (e.g., Ohala 1987, discussed in Blevins 2003: 4.5.4). On the other hand, many languages seem to take the tendency for voiced obstruents in word-final position to show reduced voicing a step further, more categorically neutralizing voicing in this position across all contexts (e.g., Kapkalli 1993 for Turkish). This distinction in pattern categoricity is sufficiently strong that it serves as one of the bases on which phoneticians and phonologists distinguish their objects of study (e.g., Hayes and Steriade 2003).

If many phonological patterns are thought to find their source in phonetic tendencies (Ohala 1971, 1975, 1981, Hayes 1999b, Bybee 2001, Blevins 2003, Hayes and Steriade 2003), how and why does increased categoricity arise in the process of 'phonologizing' a phonetically based tendency? Within formal theories of language competence, categoricity has generally been proposed to follow directly from limitations on the machinery available to the phonological grammar (discussed in section 1.5 below). In contrast, this thesis explores the hypothesis that categoricity is not directly required by the machinery of the grammatical system, but rather develops spontaneously when a phonetically derived pattern becomes sufficiently entrenched in the lexicon, where categoricity is driven by positive feedback loops reinforcing similarity in output form over many cycles of production and perception.

In chapter 1, I explore the hypothesis that markedness tendencies are not specified in the grammar, but are rather responses to grammar-external differences in the fitness[1] of particular segments, structures or sequences in

---

[1] Informally, many phonologists use the term 'marked' to refer to an element that is 'bad' in some way, whether through relative difficulty in articulation, perception, or processing. In this thesis, the term 'fitness' will be used as the obverse of 'markedness' in this sense, where fitness describes the efficiency

transmission (e.g., Ohala 1981,Lindblom et al. 1984, Lindblom 1992, 1998, reviewed in Blevins 2003). In the context of a feedback-driven grammatical system sensitive to similarity, it will be shown that differences in fitness result in categorical entrenchment of patterns over time. Furthermore, in the case of pattern conflict, it will be shown that categorical dominance of one or pattern over the other is the most stable state of the system. In particular, it will be shown that outcomes of pattern conflict predicted by the Optimality Theoretic (OT, Prince and Smolensky, 1993) principles of (i) constraint dominance, and (ii) strict constraint dominance are stable states of the system. Finally, we will see that this model predicts that even if differences in fitness are in reality additive (as seems reasonable if markedness patterns are grounded in the physical properties of articulation, perception, and processing), they may yet be manifested in lexical patterns as if they were not, as predicted by the OT principle of strict domination.

While chapter 1 investigates grammatical patterns that are grounded in fitness conflict, chapter 2 explores the hypothesis that certain patterns in

---

with which a particular feature is transmitted. Failure of a feature to be transmitted efficiently can lie at any point in the transmission pathway – for example, because it is not articulated, perceived or categorized easily. The term ‘marked’ will be reserved here for the meaning ‘crosslinguistically rare’.

lexicon structure derive from the conflicts inherent in maximizing fitness on the dimensions of articulation, perception and contrast. To do so, the simulation architecture used in Chapter 1 is modified to include a categorization step, in which greater ease of categorization increases the influence of a word on the evolution of that category. This architecture is used to explore two particular general properties of phonologies. First, phonologies exhibit only a subset of cross-linguistically attested contrasts, formed from a subset of possible features in combination. Second, crosslinguistic and intra-lexicon markedness are correlated, i.e., when a phonology allows a crosslinguistically rare element, that element tends to appear less frequently in the lexicon than the more common elements allowed by that phonology (Ferguson 1963, Greenberg 1966, Frisch 1996). Results from these simulations show that when categorization is included in the simulation architecture, outputs of the grammar become subject to selection for contrast, which in conjunction with differential fitness in transmission, results in the evolution of simulated lexical systems showing precisely these patterns.

Chapter 1 is organized as follows. Sections 1 and 2 introduce self-organization, the characteristics of systems in which it arises, and provide arguments that language in general displays such characteristics. Section 3 reviews a selection of previous work done with computer simulations suggesting that self-organization can account for certain linguistic patterns in the areas of semantics, syntax, and morphology. Section 4 returns to phonology, briefly reviewing generative accounts of categoricity in phonological patterning, followed by the proposal of an alternative source of categoricity based in the notion that outputs are under steady pressure to become more alike. General properties of such 'analogical' systems are discussed in section 5, showing how over many cycles, pattern reinforcement results in spontaneous emergence of categoricity. Sections 6 and 7 describe results from a simulation designed to test the notion that pattern categoricity, as well as categoricity in the outcomes of pattern conflict, arise spontaneously in an analogically evolving system.

1.1 Complex systems and self-organization

Self-organization serves as a cover term referring to processes in which global order operating over long temporal scales emerges through local, faster-scale interactions between system elements (Nicolas and Prigogine 1977). The well-known cellular automata program, the Game of Life, serves as a very simple example of a self-organizing system. In a typical version of this program, each cell in a grid will be 'alive' or 'dead' in a given round depending on how many adjacent cells were alive in the previous round; any cell adjacent to two or three live cells will be alive in the next round, Depending on the starting number and arrangement of living cells, these simple rules can produce a very large range of self-propagating behavioral patterns.

In general, interesting kinds of self-organization arise when the following conditions are met:

1) System elements are continually or repeatedly subject to conflicting forces.
2) Interactions between system elements are non-identical.
3) The system does not reach equilibrium on the time-scale of local change.

These conditions are quite general, with the result that any system of sufficient complexity is likely to fulfill these criteria in one way or another. In fact, as work in this field has developed over the last 30 years, it is becoming increasingly clear that self-organizational behavior is ubiquitous (Kauffmann 1993).

As a concrete example, consider the emergence of the patterned ground found in subarctic regions around the world, illustrated on the cover of the January 17 2003 issue of the journal Science (Kessler and Werner 2003). In this phenomenon, surface stones in regions subject to repeated freeeze-thaw cycles often become arranged in large circles or polygons. The environment of surface stones in sub-arctic regions fulfill conditions 1-3 above in the following way:

1) Multiple forces repeatedly interact: The cyclic expansion and contraction of water as it freezes and thaws at different rates moves stones and soil particles in opposite directions, resulting in the creation of stone piles. Gravity in turn acts against the raising up of stone piles.
2) The effect of these forces on an individual stone is highly dependent on features of the local environment, e.g., discontinuities in the proportion of rock to soil, and the local slope of the growing mound.
3) The system is not in equilibrium; for example, the forces of wind and rain do not completely randomize the mixture of stones and soil between freeze-thaw cycles.

Conditions (1) and (2) together provide the mechanism for development of local heterogeneities, in this example, local concentration and upheaval of stones. Satisfaction of condition (3) allows the effects of previous freeze-thaw cycles to persist long enough to influence the outcome of a subsequent cycle.

To see how this works, imagine that we begin with an essentially featureless soil surface. Any random irregularities in the distribution of stones and soil will affect the distribution of water and the rate at which it freezes. During a subsequent freeze, stones are pushed a little bit toward stone-rich regions, and soil toward soil rich regions, amplifying the original irregularities. Condition (3) – that the effects of the previous freeze-cycle persist into the next – allows positive feedback cycling, which is a prerequisite for self-organization. As the system evolves, the small random differences in stone-soil distribution in the original surface continue to be amplified. Eventually, derived surface features come to alter the continued evolution of the landscape: when growing dips and mounds bump into one another, their interactions under the influence of water flow, freezing rate and gravity provide new pathways for change not present in the original landscape. The eventual result is a landscape of soil ridges in geometric patterns on a dramatically larger scale than the original dips and mounds.

There are three features of this mechanism of pattern formation that are instructive in this context. First, structure formation begins with amplification of what may be undetectably small, random irregularities. Second, the mature pattern is the cumulative result of interactions on more than one level of structure. Stones and soil interact to form larger patterns of local dips and mounds, which then themselves begin to interact to form regular webs of geometric patterns of ridges. The emergence of structures at distinct scales, with distinct modes of interaction, is a general feature of complex, self-organizing systems (Nicolas and Prigogine 1977, Kauffmann 1993). Third, Kessler and Werner show in their simulations that smooth changes in the proportion of stones to soil can lead to abrupt changes in the pattern of stone organization, matching the patterns observed at actual sites. Non-linear dependence of global patterns on input parameters is a general feature of systems that include positive feedback loops (Cooper 1999).

These three properties of self-organizing systems –

i. The amplification of very small starting differences through positive feedback;

ii. The emergence of new, interacting structures as a result of interactions at smaller scales;

iii. The potential for abrupt changes in global structure accompanying small changes in local parameters;

– can make it difficult or impossible to predict the properties of the system as a whole from simple inspection of starting conditions. Because of the enormous number of possible pathways that such a system can take, computer models are useful in making and testing hypotheses about the processes driving particular dynamical systems (Cangelosi and Parisi 2002, Turner 2002). Under the hypothesis that some pattern found in language is emergent, simulation represents an appropriate strategy for developing and testing hypotheses about the properties of, and interactions between, smaller-scale elements proposed to drive the emergence of that linguistic pattern. In turn, if we find that such simulations evolve global properties analogous to those found in language, the hypothesis that such global properties could be emergent rather than pre-specified is supported.

1.2 Language as a self-organizing system.

A number of researchers have argued that language exhibits self-organized structure at a number of levels (e.g., Lindblom 1986, Labov 1994, Cziko 1995, Hurford 1999, Cooper 1999), and in fact, it seems clear that language is likely to fulfil the criteria for self-organization laid out above in a number of ways. These criteria are repeated below, followed by a summary of proposed influences on language that match the criteria.

1) System elements are continually or repeatedly subject to conflicting forces.

Many patterns in language have been proposed to be directly or indirectly influenced by the conflict between multiple influences on output form. Within phonology for example, the notion that conflict between minimization of articulatory effort and maximization of perceptual distinctiveness has an influence on grammatical patterns has held currency at least since Baudoin de Courtenay (1895[1972]). Contemporary work grounding phonological patterns in optimization of conflicting influences on output form include work done within Natural Phonology (Donegan and Stampe 1979, Stampe 1972), Grounded Phonology (Archangeli and Pulleyblank 1994), and Optimality Theory (Prince and Smolensky 1993) to name a few. More recently, patterns in the forms of individual lexical items have been explained through the conflict between minimization of articulatory effort and maximization of lexical access efficiency independent of auditory perception *per se*, at both phonological (Wedel 2002, Ussishkin and Wedel 2002) and phonetic levels (Wright 1996, Scarborough 2003, Wright (forthcoming)).

2) Interactions between system elements are non-identical.

Within phonology, a wide variety of kinds of distinctions in the relationships between elements have been proposed to underlie pattern differences. One fundamental distinction has been proposed to lie in differences in confusability between differently categorized sounds. For example, the observation that word-final [d] is more likely to be misheard as [t] than as, say [m], has been proposed to underlie the relative frequency of phonological processes neutralizing underlying distinctions between /d/ and /t/ word-finally, and the corresponding absence of processes that neutralize the distinction between underlying /d/ and /m/ (e.g., Ohala 1981, Steriade 2001, Blevins 2003).

Further, differences in confusability at the word, as opposed to phoneme or feature level, have been proposed to underlie a variety of phenomena, from word-class based contrast neutralizations (e.g., Ussishkin and Wedel 2002) to paradigm-based anti-homophony processes (Blevins 2003). In fact, nearly any work in phonology directly or indirectly supports the claim that similar phonological categories (whether that similarity is stated directly in terms of perception or not) interact more readily than more dissimilar categories (e.g., Lindblom 1998, Flemming 1995, Burzio 2001, Ni Chiosain and Padgett 2001, Pierrehumbert 2001, Padgett 2002, Blevins 2003).

Beyond their effects on perception, differences in adjacency relations within individual strings can play a role in the development of phonological patterns. Crosslinguistically, differences in adjacency are strongly correlated with the probability and nature of particular assimilatory and dissimilatory processes, e.g., locality and blocking in vowel harmony (Archangeli and Pulleyblank 1989, Ni Chiosain and Padgett 2001), and distance dependence in the violability of the Obligatory Contour Principle (Frisch 1996).

This is a non-exhaustive summary, to say the least. What should be clear however, is that for any given proposed interaction between phonological elements, we frequently can identify differences in the nature or extent of that interaction based on the identity and context of the elements in question.

3) The system does not reach equilibrium on the time-scale of local change.

The time-scale of language change within a community is much smaller than the time-scale of the actual events that are proposed to underlie that change (e.g., Bybee 1985, 2001 Pierrehumbert 2001, Blevins 2003). Every time we produce or process language, myriad individual conflict outcomes are determined, but nonetheless, the adult lexicon and grammar of a speaker does not evolve dramatically during his or her lifetime. The fact that adult grammars do not change rapidly, in conjunction with the fact that children eventually acquire a language very similar to that of community adults, regardless of the results of their initial forays into language, means that the present state of any language will exhibit strong *hysteresis*, i.e., a strong dependence on prior states. For example, despite the general tendency of language users to regularize morphophonological relations, fossilized remnants of historical phonological alternations can persist in a language for extremely long periods of time after that alternation has ceased to be productive.

These sets of general observations suggest that language presents rich opportunities for patterns to arise through self-organizational processes.

Above, I summarized some evidence that a wide variety of conflicting constraints at many levels can influence the form and understanding of utterances, and further, that various elements of language systems, whether feature and phoneme categories or lexical entries, interact in individualized ways. Finally, the inertia exhibited by language systems allows the results of previous conflict resolutions to persist long enough to influence the environment in which conflicts are resolved in the present. The resulting cyclic compounding of the outcomes of local interactions constitutes a feedback loop, which is the *sine qua non* of self-organization.

1.3 Previous work simulating linguistic self-organization

A large body of simulation studies ranging over the evolution of semantics (e.g., Oliphant 2002, Steels and Kaplan 2002, Kirby 2000, Kirby and Hurford 2002), semantics-syntax mappings (e.g., Steels 1998, Kirby 2000, Kirby and Hurford 2002, Batali 2002,), morphology (e.g., Hare and Elman 1995, Batali 2002), syllable structure (Redford et al. 2001) and vowel-systems (Joanisse and Seidenberg, 1997, de Boer 2000) demonstrate that cycling of simulated language systems results in self-organization, in which larger-scale, linguistically relevant patterns spontaneously emerge from specified, smaller-scale interactions (Pierrehumbert 2001a,b, 2003). To provide a closer look at how propagation of systems of interacting elements can self-organize to produce patterns found in language, I provide below a sketch of Hare and Elman's 1995 simulation of the morphological evolution of Old English verb paradigms.

Hare and Elman's morphological study used iterated, error-prone learning by a sequence of neural networks to model the evolution of the various forms of the English past tense in the transition between two stages of Old English. This transition was preceded by a series of sound changes which altered the similarity relationships between verbs within and across paradigms. In this simulation, a neural net with limited computational resources was given the task of learning the correct present-, past-tense pairings of a large set of Old English verb forms as they existed in the period just after the sound-changes had occurred. The limitation in computational resources deprived the network of sufficient resources to learn all forms by rote with 100% accuracy, which had the following crucial consequence: because the neural net could rely on generalization to guess a correct past tense form if rote memorization failed, if a given verb's correct present-past mapping was parallel to other similar verbs' mappings, it would have a greater chance of being reproduced accurately. On the other hand, if a given verb's present-past mapping was idiosyncratic, it would have a lower probability of being reproduced accurately.

As a consequence, although the network learned the correct past tenses for nearly all the verbs, those that it learned incorrectly were predictably those verbs that were highly irregular given the sound changes that had occurred and/or were particularly infrequent, such that the network had fewer opportunities to learn their mappings. Verbs were usually mislearned through inferring a regular past tense rather than the presented irregular pattern. As predicted however, the network occasionally also mislearned a regular verb if it was sufficiently similar to a very frequent irregular verb, or to a well-represented group of less frequent irregular verbs.

To model the evolution of English over many generations, Hare and Elman presented the output of one network, complete with its small number of incorrectly-learned present-past mappings, to a new network. The output of this second network was then passed on to a third, and so on. This process of repeated transfer of knowledge has been termed *iterated learning* (Kirby 2000). Each network learned the pattern of the previous network nearly 100% correctly, but because the errors of each network were passed on to the next, errors were able to accumulate such that after a number of generations, the pattern of regularity and irregularity had shifted significantly from its starting point. Supporting Hare and Elman's contention that iterated learning under analogical pressure can account for attested patterns in morpho-phonological change, the shifts that did occur in each run of the simulation, though different each time in their details, paralleled the historical changes that occurred in Old English.

A parallel example of an iterated learning process is the Telephone Game, in which the first in a line of people whispers something into their neighbor's ear, who whispers what they understood to the next person, and so on down the line. Although each person may pass on something very close to what the previous person whispered to them, the accumulation of errors through iterated passage usually results in a final result that is quite different than the initial utterance. It is this property that makes the Telephone Game so entertaining: the result is different from the original utterance in a way that is hard to predict, and furthermore could not plausibly have arisen through one transfer, no matter how poorly articulated the original utterance. This is because each error introduced in one round brings new features to the sentence that cannot be distinguished by the hearer from the original 'correct' features. This property, that errors participate fully in influencing the range of possible interpretations in a subsequent round, allows the sentence to rapidly and unpredictably evolve into new territory. In contrast, in the absence of iteration, all that can be modulated is the intelligibility of the initial set of words, with the result that a correspondingly smaller potential range of plausible interpretations is available. In general, whenever cyclic transmission of a pattern proceeds via passage through an information bottleneck followed by reconstruction, as in the Telephone Game, and Hare and Elman's simulation of English past-tense evolution, patterns arise that cannot be accounted for solely by reference to the starting point (Nicolas and Prigogine 1977, Kirby 1999, 2000; Kirby and Hurford 2002). Kirby (2000) has argued cogently that the bottleneck provided by the reconstruction of an individual's lexicon and grammar (I-language, Chomsky 1986) from the patterns present in the community (E-language, ibid.) provide much of the raw material for the self-organization of grammatical systems over time.

1.4 Categorical behavior in phonology

One of the most salient properties of phonological systems is their very consistency. In American English for example, if an underlying /t/ is in the appropriate environment for flapping, it will surface as a flap with few exceptions (Hammond 1999). Furthermore, when a string satisfies the structural description of multiple conflicting generalizations active in a phonological system, the outcome of that conflict shows a strong tendency to be consistent from one case to the next; in Optimality Theory, this tendency is accommodated by the principle of constraint domination. For example, Afar exhibits both a constraint against geminate consonants, and a rule of syncope, which conflict when syncope would produce a geminate (Bliese 1981). The outcome is a consistent failure to syncopate just in case a geminate would result.

The work presented in this chapter suggests that this kind of categorical behavior in phonological systems can be accounted for as an emergent property resulting from self-organization. To provide a framework for later discussion, I briefly review below accounts of phonological categoricity in current theoretical systems.

1.5 Generative accounts of categorical behavior in phonology

Rule systems (e.g., Chomsky and Halle 1968) account for consistent patterns in surface form in given contexts through context-sensitive rewrite rules that act upon underlying forms stored in the mental lexicon. Categorical behavior follows from two assumptions: 1) that rules manipulate a small number of discrete, symbolic units that correspond to categorical patterns in production, and 2) that if the structural conditions of the relevant rule are met, a rule must apply, otherwise it must not. Further, if rules can possibly conflict, they are ordered with respect to one another, allowing one of the rules to consistently determine the conflict outcome. The observation that rule-conflict outcomes are consistent across a language is accommodated by stating that within a grammar, rule orderings are invariant,. Ordering of rules is language-specific, so that when we find that in a related language it is the other rule that determines the output in case of rule conflict, we say that the rule ordering is reversed[2].

Whereas rules systems rely on application of inviolable rules to derive input-output relations, in Optimality Theory (OT, Prince and Smolensky 1993) lexical inputs are mapped to optimal outputs through the satisfaction of violable, ranked constraints on output form. The set of constraints is universal, but the ranking of those constraints is language specific, accounting for language-specific output patterns. Where categorical behavior in rule-systems is achieved through the specifications and orderings of the particular rules operative in a given grammar, output patterns are categorical in OT because

---

[2] Conflict avoidance can also be achieved by simply building it into the rule itself. In the Afar example above, a syncope rule could specify that vowels delete between consonants, provided those consonants are not identical.

the ranking of constraints is fixed, and therefore any input strings that share a relevant set of morpho-phonological properties will exhibit an analogous mapping relationship to their optimal outputs. The ranking, or *dominance* relation between two constraints is made apparent when an optimal output violates one constraint in order to satisfy the other. For the purposes of discussion here, I will use the term 'dominance' beyond its strict OT usage in reference to the relation of OT constraints to refer to the more general categorical satisfaction of one surface pattern at the expense of another within a language.

In addition to simple dominance relations between conflicting patterns, grammars often exhibit a higher-order kind of dominance that becomes apparent when multiple patterns collide in one output form, in which the result of a conflict between multiple patterns tends to follow the results of the component pair-wise pattern conflicts. More concretely, if pattern A wins out over both patterns B and C separately, pattern A will still win out to produce a form that violates both B and C together, even if there is an alternative that violates A while simultaneously satisfying B and C. OT accommodates this observation through the stipulation that constraint dominance is *strict*, that is, that output candidates satisfying a higher ranked constraint will always win over candidates violating that constraint, while satisfying any number of lower ranked constraints. Stated less formally, lower ranked constraints cannot 'cooperate' to compel violation of a higher-ranked constraint. Again, for the purposes of discussion I will generalize the OT term 'strict dominance' to refer to the persistence of pattern dominance in the face of conflict with multiple, agreeing patterns.

1.6 Framework of the simulation

An aim of this chapter is to show that pattern relations exhibiting both dominance and strict dominance emerge spontaneously within simulations grounded in psycholinguistically supported models of the lexicon. The simulation architecture used here is based in a general model of language production and processing that satisfies the following two general conditions:

1) In addition to storing more abstract categorial generalizations, the lexicon is able to store sub-phonemic information influenced by individual events (Goldinger 1996, reviewed in Tenpenny 1995 and Johnson 1997).

Since Baudoin de Courtenay (1895 [1972]), linguists have proposed that factors in the linguistic environment give rise to directional, gradient biases in the form of utterances, which ultimately serve as the raw material for phonologization (see e.g., Ohala 1981, 1989, Bybee 1985, Kiparsky 1995, Blevins 2003, for more recent arguments in favor of this position). In this model, storage of phonetic detail provides a mechanism for low level drift in phonetic behavior to feed back to the lexicon, providing the raw material for (occasionally abrupt) larger-scale shifts in phonological behavior (Cooper 1999).

2) The mechanism for assembling production targets for a given linguistic element allows such targets to be biased toward the form of other, similar linguistic elements.

Clustering of elements into groups on the basis of features held in common is a recurrent feature of language, in the realms of morpho-phonology (e.g., Plunkett and Marchman 1991, Burzio 2002, Albright 2002, Krott, Baayen and Schreuder 2001, reviewed in Bybee 2001), phonology (reviewed in Burzio 2001, 2002; Bybee 2001), as well as in the organization of articulatory motor sequences (Browman and Goldstein 1988, 1990, Saltzman and Munhall 1989, Byrd 1995, Bybee 2001) and motor sequences in general (Shadmehr and Bashers-Krug 1997). Biasing of outputs towards previous behavior will be shown to create a feedback loop that steadily works to level out distinctions in behavior in the lexicon. This leveling behavior will be shown to result in the development of categorical behavior.

The simulation architecture rests on three conceptual elements:

i. a lexicon consisting of lexical entries, each comprising an abstract ‘phonemic’ level of representation and an associated, more phonetically detailed set of representations;

ii. an implementation mechanism that uses the information inherent in the lexicon to produce production targets, where lexical outputs are biased toward other forms in proportion to similarity;

iii. a performance filter that introduces directional biases in actual output form.

Within each round of the simulation, each lexical category produces a set of corresponding outputs, where output shape is based not only on information stored in that category, but also on information stored in other categories, in proportion to similarity. Once an output is assembled, it is passed on to phonetic implementation, where performance biases introduce a small amount of directional noise in the final output form. Production is followed by re-storage of the output in the source category. Note that because outputs are automatically reassigned to their source category, this architecture does not model categorization, but only the feedback-driven evolution of lexical patterns under the joint influences of a bias toward similarity of output form from within the lexicon, and external performance biases in production. An elaboration of the model including a categorization step will be introduced in Chapter 2.

Because each category is maximally similar to itself, the greatest influence on output form comes from the source category itself, with the result that output forms tend to closely resemble previous outputs of the source category. However, in the event that a given pattern in the lexicon becomes sufficiently common across lexical categories, it will begin to have a more significant effect on production across the lexicon. Because of the positive feedback inherent in the production-storage loop, if a pattern does happen to gain a foothold in some part of the simulated lexicon, it will tend to 'snowball', rapidly spreading across the entire lexicon to produce a stable categorical pattern. It is this single property of the model that forms the basis for the results presented in this chapter.

Sections 1.6.1-3 discuss these three elements of the model in greater conceptual detail, along with their ramifications for pattern formation. Section

1.7 provides greater detail concerning the machinery of the simulation that models these elements and their interaction.

1.6.1 Phonetic detail and the structure of lexical entries

Many computational models of speech processing posit multiple levels of linguistic analysis (e.g., TRACE, McClelland and Elman 1988; Shortlist, Norris 1994; PARSYN Luce et al. 2000; reviewed in the context of work in theoretical linguistics in Jusczyck and Luce 2002). In support of this modeling strategy, experimental evidence from priming has been found in support of the position that both surface and underlying form levels of representation participate in speech processing (McLellan, Luce and Charles-Luce 2003).

Within the simulations presented in this chapter, each lexical category comprises two explicit levels of representation: an abstract 'phonemic' level containing a single string, and a second, more detailed level containing a number of recalled exemplars (Goldinger 1996, reviewed in Tenpenny 1995) of that category (Pierrehumbert 2001a, b, 2003). For example, given a lexical category with an abstract level composed of the string /ubli/, a set of corresponding exemplars might additionally retain information on syllabification, such as [ub.li], or [u.bli] (where a (.) represents a syllable boundary, and slashes and square brackets represent more and less abstract representations, respectively). The single lexical entry for ‘ubli’ within a simulated lexicon might therefore look like the following:

Figure 1. Sample Lexicon

/ubli/:	[ub.li]
	[ub.li]
	[ub.li]
	[u.bli]

In this example, the lexical entry for ‘ubli’ contains an abstract level /ubli/ with a set of four recalled exemplars containing in addition information on syllabification. Three of the four exemplars in this example were stored with an inter-consonantal syllable boundary, while the fourth recorded a pre-consonant cluster boundary.

An exemplar approach is used here for the most detailed level of representation, because it both provides a computationally convenient way to represent potentially heterogeneous detail within a single level of representation, and a convenient mechanism for abstracting various levels of categorial detail from a single level of representation. Most importantly in the present context, exemplar models provide a clear mechanism for bridging the gap between the phonological and the phonetic, that is, between categorical phonological behaviors, and the gradient effects that have long been hypothesized to underlie them (see Pierrehumbert 2001a, b; to appear). Additional background on exemplar models of linguistic categorization phenomena is provided in section 1.6.1.1 below.

1.6.1.1 Exemplar Theory

The assumption of a fundamental distinction between general, abstract knowledge and specific, episodic memory has a long tradition in the psychological literature on categorization. In recent decades however, research has repeatedly found that subjects retain access to highly detailed, episodic memories of an event for a surprisingly long time (reviewed in Johnson 1997), and make use of these memories when carrying out tasks thought to require only general knowledge (see Tenpenny 1995 and references therein). As a consequence, a class of new theories has developed which locate specific, episodic memories at the core of categorization processes (Hintzman 1986, reviewed in Jacoby and Brooks 1984). While such theories do not deny that generalizations exist, they begin from the hypothesis that abstract knowledge has no special status relative to specific knowledge, and that abstract knowledge does not necessarily require a form of representation distinct from that encoding specific memories. In the last decade, these so-called *exemplar* models have been extended to the domain of language by linguists and psycholinguists interested in categorization phenomena both in perception (Goldinger 1996, Johnson 1997) and in production (Goldinger 2000, Pierrehumbert 2001a, b, 2003).

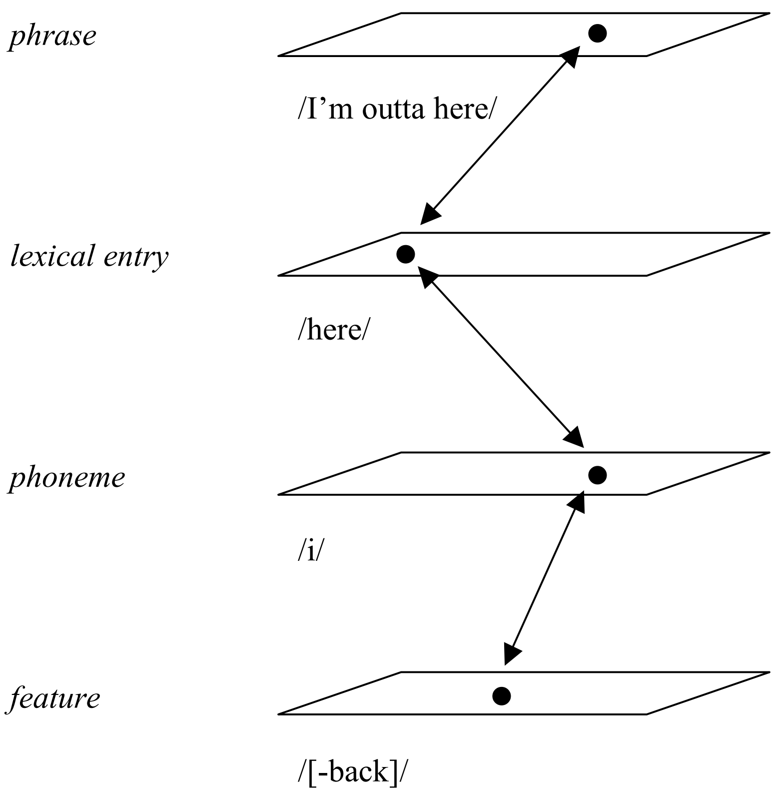
In these models, each category is defined by a 'cloud' of remembered tokens, or exemplars, that have been tagged as belonging to that category. Exemplars are organized within the category by similarity across any salient dimension, producing internal structure in category-space; a given exemplar may therefore contribute to many categories simultaneously. Depending on the model, new experiences are assigned to relevant categories by comparison to actual exemplars (reviewed in Tenpenny 1995), or to generalizations directly emerging from the exemplars that make up a category (e.g., Hintzman 1986, Goldinger 1996), and inserted into category-space at the appropriate point. Because future categorization events are carried out by reference to previously categorized exemplars, exemplars in a category necessarily display the range of characteristics displayed by physical members of that category. In the process of identification and storage, each experience changes the content of memory, either by entering a new exemplar trace, or through leaving behind a higher level of activation of an indistinguishable exemplar that was previously stored (e.g., Kruschke 1992). As a result, frequent categories are populated with a higher density of more highly activated exemplars relative to infrequent categories, accounting for a number of frequency-related phenomena, including for example, more rapid, accurate category assignment of high-frequency events (Kruschke 1992 and references therein). Note that in exemplar models, frequency is not specifically encoded in the system, but is intrinsic to the processing mechanism. These models have been argued to successfully account for a range of linguistic phenomena that have recently come to light, such as long-term priming of subphonemic phonetic detail (Goldinger 1996, reviewed in Tenpenny 1995), frequency-dependent shadowing latencies (Goldinger 1996), and word-specific allophony (Goldinger 2000, reviewed in Jurafsky, Bell and Girard 2002).

1.6.2 Output assembly within the lexicon

Within a general exemplar model of the lexicon, the lexicon may potentially contain exemplars in all linguistically relevant categories: allophones, phonemes, strings of segments, words, phrases and so on (e.g., Goldinger 1996, reviewed in Pierrehumbert 2003). In support of this general hypothesis, the finding that phoneme identification is more rapid if a phoneme is embedded in a real, as opposed to nonce word, has been interpreted to indicate that both phoneme-size and larger categories are able to additively participate in sound recognition (Ganong 1980, discussed in Burzio 2001).

In this type of model, every incoming chunk of information can potentially be broken apart and categorized at many levels, as exemplified in Figure 2 for an instance of the stock phrase, 'I'm outta here!', which can be located as a whole in the parameter space of the lexicalized phrase itself. The component word 'here' also contributes an exemplar in the parameter space for the lexical entry /here/, and likewise the vowel [i] of 'here' can contribute an exemplar to the phoneme /i/ parameter space. Finally, the perceived [-back] feature of the [i] vowel could itself contribute a [-back] exemplar to a parameter space labeled for backness (see Pierrehumbert 2001a for additional discussion).

Figure 2. An event may contribute exemplars to multiple, nested categories



Each plane represents a parameter space for a category, identified below the plane in brackets. The point in each plane represents the exemplar contributed to that parameter space in memory through processing a single example of the phrase, "I'm outta here!". The nesting of categorization events means that every exemplar in a category potentially contains, and is contained by, other linked categories.

Under this hypothesis then, because all exemplars of a given category contain and/or are contained within other exemplars of other categories, the lexicon is characterized by a dense interconnectedness between all categories, in which production of a category (e.g., a word), may proceed under the influence of categories on many distinct levels.

Although, as detailed above, exemplar models provide a convenient structure for information storage and flow within the model, it should be stressed that the exemplar approach *per se* is not critical to the function of the simulations presented here. Rather, the architecture of the simulations presented below could be couched in any model that allows both (i) phonetic detail related to individual events to influence form at some level of representation, and (ii) a mechanism for output targets to incorporate patterns across category boundaries.

1.6.2.1 Reversion to the mean in production

Following the discussion above, the model underlying the simulations used here begins with the assumption that segment-sized categories coexist with, and are cross-referenced with, categories made up of larger strings, and that the least abstract elements of each category are highly detailed. Within the simulation, production of an output begins with selection of a lexical category, followed by assembly of the target output in full phonetic detail through reference to the array of detailed exemplars stored in that category. However, assembly of a fully detailed output does not proceed solely through reference to the previous outputs of the lexical category to be produced, but also by reference to the contents of *other* lexical categories, to the extent that they share sequences with the category to be produced.

For our purposes here, the important property of the model deriving from cross-category influence in production is that a production target is based to some degree on a consensus, with a result that outputs will exhibit 'blending inheritance', and thereby tend towards 'reversion to a mean' over the lexicon. Reversion towards the mean of a set follows necessarily when a new element is produced from a consensus over multiple prior elements of the set (Abler 1997, Pierrehumbert 2001a). For example, if a production target is composed by reference to two exemplars that are distinct on some dimension, the resulting targeted value for that dimension will fall between the values of the two exemplars, in a sense 'blending' their contributions. Consequently, in a model of this type all individual dimensional targets of a production goal can only fall *within* the extremes present within the contributing category set, never outside them. This constant tendency to assemble outputs that conform to a mean over previously stored forms constitutes, in effect, a form of analogical pressure, operating in this particular case at the level of phonological production targets to reinforce existing patterns. When allowed to feed back on itself over many cycles, this analogical pressure will be shown in later sections to interact with external biases to produce familiar higher-order phonological patterns over time.

In this context, it may be useful to compare the architecture described here with that of simulations described in Pierrehumbert (2001a, b, 2003) of the evolution of individual phonetic categories within a production/perception loop. In Pierrehumbert's simulations, the assembly of a phonetic output is modeled as selection and subsequent averaging of a subset of previously stored exemplars associated with the phonetic category under production. After selection, the average is passed on to phonetic implementation, which introduces a small amount of noise in the process of output production. Production is followed by re-storage of the output in the best matching phonetic category (see Pierrehumbert 2001a for details). Pierrehumbert shows that starting from a single seed exemplar, the variance of the distribution of exemplars within a category increases rapidly at first due to stochastic changes introduced in production, but soon stabilizes, becoming entrenched around some value. However, Pierrehumbert shows that this stabilization crucially depends on the fact that an output of production does not derive from a single exemplar, but rather derives from a *consensus* over a subset of exemplars from that category. It is this blending of characteristics in production that results in steady reversion to the mean of the category in production, and eventual category entrenchment. (Pierrehumbert goes on to show that introduction of directed bias in the noise of phonetic implementation results in gradual shift of the phoneme category in the direction of the bias, modeling phonetic drift under the influence of articulatory markedness.)

For our purposes here, the relevant distinction between Pierrehumbert's simulation architecture and the one used here lies in the fact that here, the influence on the form of an output can originate in categories beyond that under production, in proportion to similarity. Because intra-category similarity will be, on average, greater than cross-category similarity, categories will still show a tendency to entrenchment over the short term. However, the existence of any degree of cross-category influence should slowly but surely result in the evolution of categories to become increasingly alike.

In the remainder of this dissertation, I will refer to systems in which patterns are propagated through this kind of similarity-based blending inheritance as analogical systems, and the result of blending inheritance will be referred to variously as analogical pressure or pattern reinforcement. Because patterns are propagated via some degree of blending inheritance, analogical systems are characterized by a steady evolution toward uniformity, in the absence of other factors introducing or maintaining difference. Further, due to the dependence of cross-category influence on degrees of similarity, the pathway to system uniformity in an analogical system is not smooth, but rather characterized by the temporary development of sub-regions of uniformity, that then spread. In the following section, I illustrate this property of analogical systems using a simple cellular automata simulation.

1.6.2.2 Properties of analogical systems

A system can be described as analogical when the future behavior of a system element is biased towards the present behavior of other system elements on the basis of some measure of similarity. There are two properties of such systems that will concern us here:

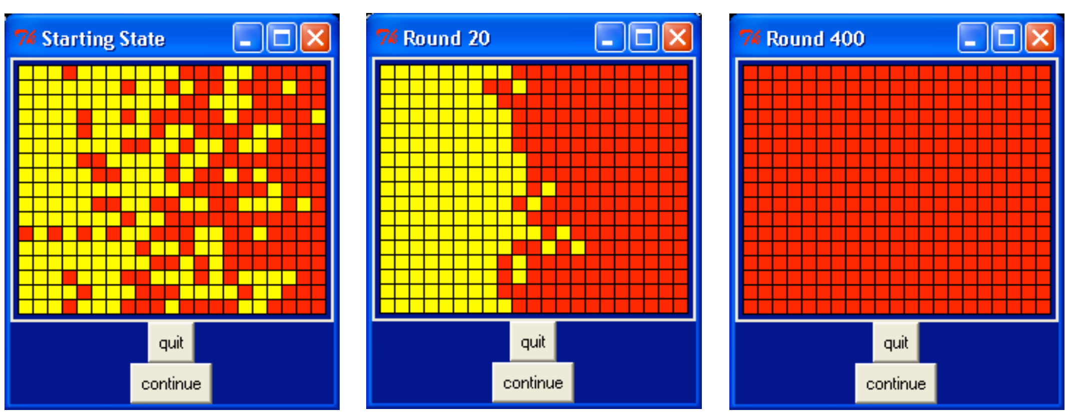
1) Gradient patterns in behavior are unstable; persistent bias towards similarity between elements promotes the development of sharp boundaries in behavior.

2) In the absence of forces that maintain or add difference in a finite system , all system elements will eventually come to exhibit identical behavior.

Both properties directly derive from the bias toward similarity. To illustrate this, I show stages in the evolution of a simple cellular automata program (Figure 3A-C).

Figure 3. Cellular Automata Program

A. B. C.



At the starting point of the simulation, each square in the field is randomly assigned the shade *dark* or *light*, with the caveat that there is a gradiently higher probability of being assigned *dark* from left to right. This starting state is shown in Figure 3A, where we see that there is indeed a gradient pattern of *light* to *dark* going to the right. In each subsequent round of the simulation, each square has a small probability of changing its shade. Crucially, however, the choice of shade is stochastically biased towards that of squares in the vicinity, with more weight given to nearer neighbors' shades. In this simulation then, distance between squares is the relevant dimension of similarity. If the shades of a square's neighbors happen to be equally distributed between dark and light, then the square will change randomly—but the more of any given shade there is in the neighborhood, the more likely it is to change to match. The result is that the initially gradient *light-dark* pattern rapidly becomes more categorical, as can be seen in Round 20 (Figure 3B).

Note that this segregation occurs despite the fact that there is no explicit mechanism in the conditions of the simulation to directly cause segregation of shades: there is only a direction to become more like one's neighbors. However, the result of this direction is that squares change their shade often when in mixed neighborhoods, but stay put when there is local shade-consensus. This difference in rate of change creates a basin of attraction formed by shade-identity, with the result that consensus neighborhoods expand at the expense of mixed neighborhoods. The greatest reduction in boundary between two neighborhoods in this 2-dimensional field is a straight line, which the simulation continues to approach beyond the point reached in Figure 3B.

This is entirely analogous to the familiar physical system of oil droplets coalescing to form a large circle on the surface of water in a pot. This effect derives from one simple fact: water molecules stick more tightly to water molecules than to molecules of oil. To see how this results in separation of oil and water, imagine we start with a well-stirred mixture of the two. As all the molecules jiggle and bump into one another, trading one molecular neighbor for another, water molecules tarry longer with other water molecules than with oil molecules. This difference in the rate of exchange of neighbors results in a steady accumulation of water-only neighborhoods, and the compensatory creation of oil-only neighborhoods. Eventually, the water and oil are entirely separated (at least to the eye) – with the oil in a beautiful circle, which is the shape that provides the minimum region of water-oil contact. Note again, there is nothing in the chemistry of water that says that oil and water molecules cannot be in contact, nor is there any mention of circles or spheres. Separation, and the resulting geometry of that separation, are driven solely by water molecules' greater stickiness for other water molecules than for oil molecules.

To see that analogical systems eventually evolve to uniformity when left alone to their own devices, return to the simulation at the stage in Figure 3B. We can see that squares at the boundary between light and dark will often switch shades, as their neighborhoods are nearly evenly matched between shades. As a consequence, the boundary between dark and light migrates randomly back and forth across the field. If at some point the boundary gets too close to one side, the weight of numbers on the majority side will increasingly influence the squares of the minority shade, accelerating the movement of the boundary towards the edge. When a boundary hits the edge, one of the shades disappears from the field – and since that shade can only arise in a square through the influence of another square of that shade, it is gone for good. This scenario came to pass in favor of *dark* in the simulation shown in Figure 3 at round 400 (Figure 3C). This result is general: in any finite system evolving solely through blending of features, progressive loss of extremes is inevitable, until the system becomes uniform (Abler 1997).

For our purposes here, the crucial property of an analogically evolving system is the following: on the pathway from complete disorder to complete uniformity, the system passes through stages in which distinct patterns become segregated into sharply bounded regions of similarity. Therefore, in a system that is prevented from reaching uniformity through external re-introduction and/or maintenance of difference, we expect to find system elements segregated into defined regions of similarity, whose boundaries shift slowly with time.

1.6.3 External biases on output form

Within the model under consideration, cross-category influence in production steadily biases output forms to become more alike. But no lexicon comes to eventually consist of one word – on the contrary, diachronic change gives the impression of a constantly shifting equilibrium. Within this model, such an equilibrium can only be maintained through forces that support or introduce difference within the lexicon. In Chapter 2, we will see that competition between categories in lexical access (Luce and Pisoni 1998) can account for maintenance of contrast in evolving lexicons. In this chapter, we'll investigate a distinct, contributing source of variation found in context-sensitive biases in performance –those gradient, phonetic-level markedness tendencies that linguists from Baudoin de Courtenay (1895 [1972]) to Blevins (2003) have proposed form the raw material for grammaticalization. These tendencies are modeled as biases external to the lexicon, biasing output form after the output target has been assembled from patterns within the lexicon. Re-storage of final output forms in the lexicon results in a feedback loop, continually biasing the lexicon with new exemplars that bear the traces of these performance biases. We might predict then that in the ensuing shifting equilibrium, performance bias-derived gradient patterns may occasionally be converted into categorical ones within the lexicon, just as we saw in the distribution of white and black squares in Figure 3.

It has been cogently argued that biases in performance may be grounded directly in the many layers of context of use (e.g., Lindblom 1994, Bybee 2001, Blevins 2003 and references therein), implying that biases will not only be gradient, but also sensitive to any factor in the immediate environment that could possibly affect performance. This might include, for example, physical factors such as the particular shape of someone's mouth, or whether or not they are currently eating; discourse factors such as the stability of the current common ground, or the ambient noise level; or social factors such as the degree of familiarity of the setting, or in-group/out-group status. Given the enormous number of factors possibly impacting the form of an utterance at any given moment, the exact direction and degree of bias on a given output might be considerably different in any given instance. Bearing this in mind, however, because the simulation is not explicitly intended to explore the ramifications of context-driven bias variability, biases are idealized as constants in the simulations below for the sake of computational simplicity.

In the simulations presented in this chapter, complete collapse of all difference into uniformity is artificially prevented by holding the abstract, phoneme-level representations fixed, and only allowing the mapping between abstract-level and exemplar-level strings to vary. This will allow us, in addition, to specify what kinds of pattern conflict we wish to investigate. In chapter 2, a more significant source of difference – selection pressure through repeated categorization resulting in preservation of contrast– will be included in the simulation architecture, at which point lexical entries will be allowed greater freedom to evolve. The immediately following sections describe the machinery of the simulation in greater detail.

1.7 Architecture of the simulation

1.7.1 The structure of the lexicon

Lexical entries are split into two levels, the first an ‘underlying form’ composed of an ordered string of univalent feature tags, and the second a set of more phonetically detailed, stored exemplars of previous outputs from that lexical entry. In the simulation, the content of exemplars is expressed using the same set of feature tags as the underlying forms. The simulation uses ‘features’ solely as a mechanism to compute the degree of similarity between segments; the more features shared, the more similar. This is in principle similar to feature-counting approaches to predicting similarity between segment types (e.g., Tversky 1977). However, in this simulation, feature tags are also used to record additional characteristics of a segment in a stored exemplar, such as whether that segment is in onset or coda position, or at a morphological boundary.

Note that the description of simulation elements in terms of ‘words’, ‘segments’, and ‘features’ is metaphorical – these are simply convenient names given to elements within the simulation that are related through particular set-subset relations. Hence, results from these simulations are not intended to make any theoretical claim about this or that featural theory. In fact, given that propagation through any kind of blending inheritance must influence a system to evolve toward categoricity, at this level of abstraction this simulation architecture can serve as a predictive model for language, provided some mechanism does indeed exist allowing similarity in the lexicon to influence lexical production. The features employed in the simulation, and the combinations used in the simulation are given in Figure 4 below.

Figure 4. Features and feature combinations used in the simulation.

Feature tags:

C	*consonant*
lab	*labial*
cor	*coronal*
dors	*dorsal*
voi	*voiced*
novoi	*voiceless*
liq	*liquid*
lat	*lateral*
rho	*rhotic*
V	*vowel*
hi	*high*
lo	*low*
bk	*back*
ft	*front*

Feature tag combinations:

b	C, lab, voi
p	C, lab, novoi
d	C, cor, voi
t	C, cor, novoi
g	C, dors, voi
k	C, dors, novoi
l	C, liq, lat
r	C, liq, rho
i	V, hi, ft
a	V, lo, bk
u	V, hi, bk

Features are organized into hierarchical classes such that, for example, a segment with a C feature may have one place feature from the set {lab, cor, dors} and one voicing feature from the set {voi, novoi}, while a segment with a V feature may have one backness feature from the set {bk, ft} and one height feature from the set {hi, lo}. In addition, the feature tags *onset*, *coda*, and *edge* are employed within stored exemplars to convey that a segment is an onset, coda, or at a morpheme edge, respectively; *onset* and *coda* are treated as members of the same class.

Figure 5 shows an example of a small lexicon containing two lexical entries *ubli* and *igra*, where each entry has three associated exemplars, differing in their perceived syllabification as shown by the position of a demarcating period (.). Each segment is short-hand for a set of features that together unambiguously identify the segment; in exemplars, syllabification and position at a morpheme edge is also included in the featural representation.

Figure 5. Another sample lexicon

abstract entry	exemplars
/ubli/	[u.bli]
	[ub.li]
	[u.bli]
/igra/	[ig.ra]
	[ig.ra]
	[i.gra]

To begin a simulation, a lexicon is provided, seeded with exemplars. In this example, the syllabification of the seed exemplars is random. In the simulation model presented in this chapter, abstract lexical entries are fixed, while exemplars are free to vary within limits under the influence of performance biases and pressure from other exemplars. The restriction on change within the abstract lexical entry and exemplar cloud is necessary at this point because there is nothing within the simulation architecture to maintain contrast between lexical categories. Hence, if lexical entries were not fixed, lexical entries would rapidly collapse toward uniformity under pressure from pattern reinforcement in production. However, restricting the range within which lexical categories can evolve has the following advantage that will be exploited in this chapter: by restricting pathways of change, particular conflicts can be created specifically to investigate the development of pattern dominance and strict pattern dominance. The simulation architecture presented in chapter 2 on the other hand differs by including a mechanism for contrast maintenance, allowing lexical patterns relating to the interaction of Markedness and Faithfulness to be modeled.

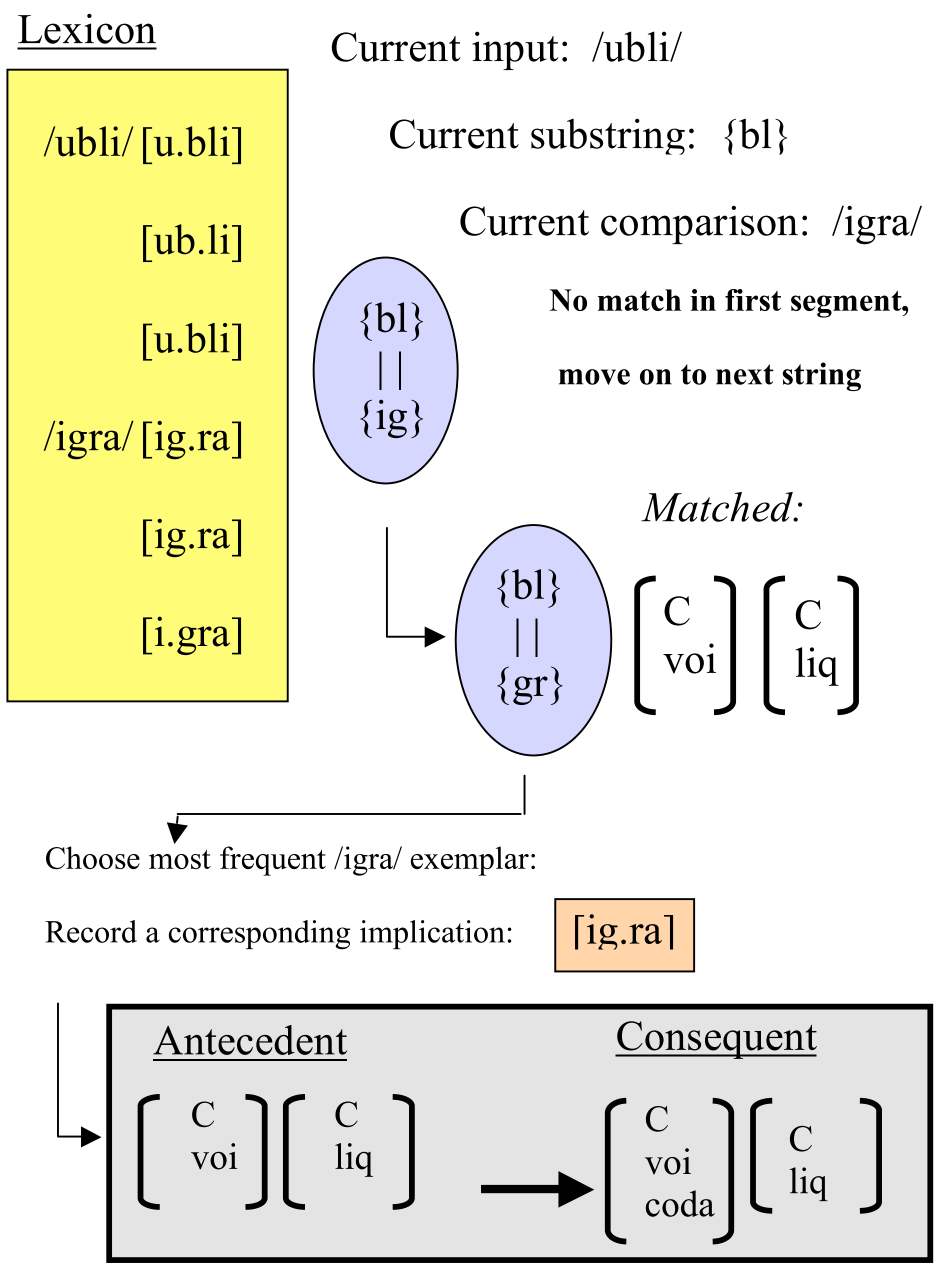
1.7.2 Simulating similarity-based connections in the lexicon

In the general model described in section 6, assembly of a production target for a lexical entry proceeds under weighted influence from exemplars of sequence-similar lexical entries across the lexicon. Likewise in the simulation, a production target for a given lexical entry is assembled through reference to both the stored exemplars from that lexical entry itself, and through the relationship between other lexical entries and their exemplars, to the extent that they share sequences with the lexical entry in question.

To simulate a web of lexical entries interacting in the assembly of a given lexical entry's production target, all lexical entry substrings that show any contiguous featural match to substrings in the given entry are identified. The set of relations between those substrings and their reflexes in corresponding exemplars are then summed to produce a probability that a given phonological category in the lexical entry will be associated with a particular reflex in the production target; a representative portion of that process in the assembly of a production target for the entry *ubli* is shown as an example in Figure 6.

First, all substrings in the lexicon that have any featural match in each segment to a given substring in the underlying entry *ubli* are identified. Substrings are defined at the segmental level, such that all single-segment substrings in *ubli* will be compared to all single segment substrings in the lexicon, followed by all two-segment substrings from *ubli* to all two-segment substrings in the lexicon, then all three-segment substrings, etc. Figure 6 illustrates the attempted match between the two-segment substring {bl} from /ubli/ and the first two-segment substring, {ig}, from /igra/. There is no featural match between {b} and {i}, so the simulation moves on to the next possible two-segment substring in /igra/, in this case {gr}. In this case, there are in fact features

Figure 6. Finding implications within the lexicon.



held in common both between {b} and {g}, and between {l} and {r}. These features are written as the antecedent of an implication. To write a consequent, the most frequent exemplar-type of the entry /igra/ is chosen, in this example [ig.ra], and the exemplar-correspondents of the matched features are written as the consequent, including any other associated phonetic detail – in this example the fact that the first segment of the exemplar substring is a coda.

In addition to implications derived from comparison between the given lexical entry and other entries in the lexicon, all implications derived from the given entry in comparison with itself are also identified.

1.7.3 Assembling a production target

After all implications pertaining to the lexical entry in production have been identified, those implications are used to construct a set of possible production targets. Continuing our example in Figure 7, the antecedent features in each implication are matched to the correct position in the lexical entry /ubli/, and the consequent features are added to cumulative lists in a corresponding feature grid.

Figure 7. Assembling a production target

- Set up an empty segment/feature grid matching the input:

/u b l i/ […][…][…][…]

- Match each implication antecedent to the input.

<u>Antecedent</u> <u>Consequent</u>

$$\begin{bmatrix} \text{C} \\ \text{voi} \end{bmatrix}\begin{bmatrix} \text{C} \\ \text{liq} \end{bmatrix} \rightarrow \begin{bmatrix} \text{C} \\ \text{voi} \\ \text{coda} \end{bmatrix}\begin{bmatrix} \text{C} \\ \text{liq} \end{bmatrix}$$

Match:

/u b l i/ → […][…][…][…]

$$\begin{bmatrix} \text{C} \\ \text{voi} \end{bmatrix}\begin{bmatrix} \text{C} \\ \text{liq} \end{bmatrix} \qquad \begin{bmatrix} \text{C} \\ \text{voi} \\ \text{coda} \end{bmatrix}\begin{bmatrix} \text{C} \\ \text{liq} \end{bmatrix}$$

After all implications have contributed their features, the result is an ordered list of segment positions, where each feature node is occupied by a list of features of that class contributed by implications containing that feature, illustrated in Figure 8. A production target is assembled by randomly choosing one feature from the list at each node. In this particular example, all the lists in all feature nodes contain only one feature specification, with the exception of the [onset/coda] node, where there are 6 votes for [onset], and 6 votes for [coda]. Hence, the randomly chosen features are 50% likely to specify the production target [u.bli], and 50% likely to specify the target [ub.li]. Once assembled, the production target is sent to Performance.

Figure 8. Selecting features.

- After all implications have contributed target features, a feature is chosen at random from the list at each slot.

/ubli/	→	[…]	[…]	[…]	[…]
		V (8) hi (8) bk (8)	C (12) lab (6) voi (12) coda (6)/ onset (6)	C (12) liq (12) lat (6)	V (8) hi (4) ft (4)

- In this example, there is an even chance of choosing features resulting in [u.bli] or [ub.li]

1.7.4 The Performance Filter

Performance contains a set biases comprising feature specifications identifying less-fit sequences, a possible change, and the likelihood of that change occurring. A bias against complex onsets, termed NoComplex[3], is illustrated in Figure 9.

Figure 9. NoComplex in Performance

$$\underset{\substack{\big\downarrow\\ \text{coda}}}{\begin{bmatrix}\text{onset}\end{bmatrix}}\begin{bmatrix}\;\text{C}\;\end{bmatrix}\qquad \text{Probability: 0.1}$$

This bias results in a 10% chance that a production target containing the feature sequence [onset], [C], will actually surface as [coda], [C]. For example, if the production target constructed by the lexicon is [u.bli], there is

---

[3] The names of performance biases within in these simulations, as well as their presumed correlates in actual performance, will be set in normal script rather than SMALLCAPS, to distinguish them from OT constraints.

a 10% chance that it will be articulated as [ub.li] instead. The output is stored in a temporary buffer, and the lexicon begins the process again for the next lexical entry. Biases in this program function only to increase the rate of change away from a given state, not explicitly toward any other state. Because in this case there is only one alternative to the feature *onset*, that is, *coda*, this bias appears to function as a directed repair mechanism, but this would not be the case if there were multiple possible changes from *onset*. When multiple biases are included in a simulation, biases apply to the output in random order.

1.7.5 Updating the lexicon

In a given round, as many outputs are produced from each lexical entry as there are exemplars, each stored in the temporary buffer. After every entry in the lexicon has produced all of its requisite outputs, the exemplars for each entry are overwritten by the corresponding outputs from the temporary buffer. The process then begins again, with new exemplars for each entry. This replacement of exemplars in each round roughly simulates the decay of exemplar memory with time (e.g., Pierrehumbert 2001a, Johnson 1997), and allows the lexicon to evolve at a constant base rate.

1.7.6 Summary of simulation architecture

The simulation fulfils the two conditions laid out in section 4, repeated here for convenience:

1) In addition to storing more abstract categorial generalizations, the lexicon is able to store sub-phonemic information influenced by individual events

2) The mechanism for assembling production targets for a given linguistic element allows such targets to be biased toward the form of other, similar linguistic elements.

Within the simulation lexicon, recalled production events from the previous round are stored, with their associated individual detail, within a labeled lexical entry. Assembly of a production target for any lexical entry proceeds using the relationship between the lexical entry labels and their associated exemplars from the entire lexicon, resulting in reversion toward the mean for the lexicon, or put alternatively, analogical pressure for similar forms to become yet more similar.

Passage of production targets through an external performance filter introduces contextually biased, stochastic changes in outputs. Re-storage of these outputs in the lexicon as lexical exemplars introduces contextually consistent biases into the lexicon, which can serve as the raw material for developing categorical behavior. The next section illustrates this process of 'phonologization' of bias in this simulation through the leveling effects of analogy.

1.8 The development of pattern domination

1.8.1 Pattern reinforcement results in categorical behavior

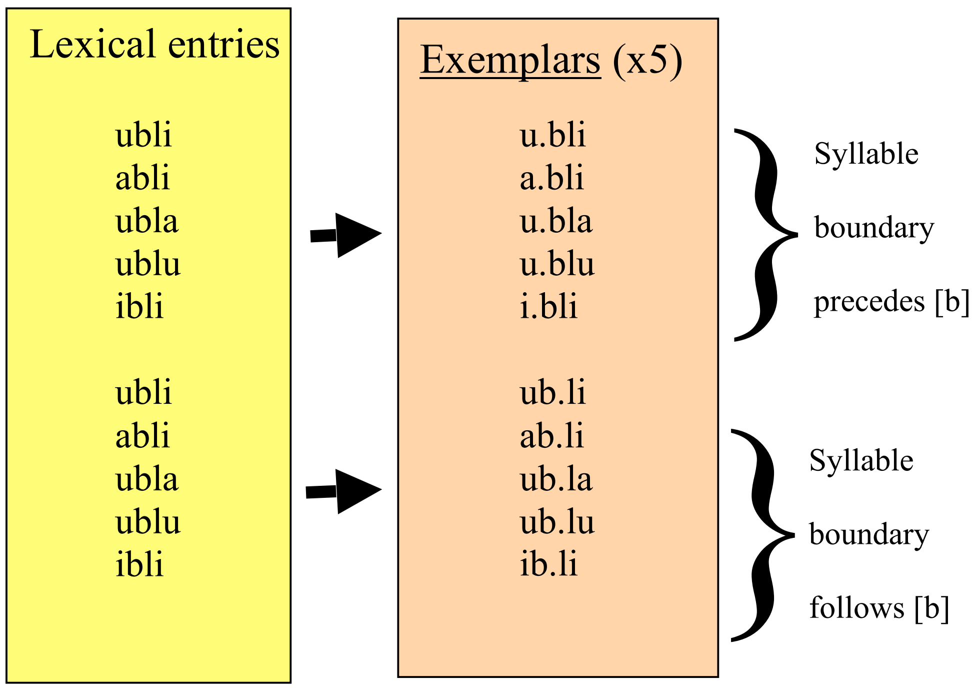
The first simulation shown uses a lexicon and seed exemplars shown in Figure 10, and uses no performance biases to filter outputs, i.e. production targets are produced without modification. Note that the list of lexical entries is doubled[4], and that the first list is exclusively associated with exemplars syllabified with internal complex onsets, while the second set is associated with exemplars syllabified with internal codas. This perfect symmetry means that in the initial

---

[4] The simulation treats all lexical entries as separate, no matter how much phonological content is shared.

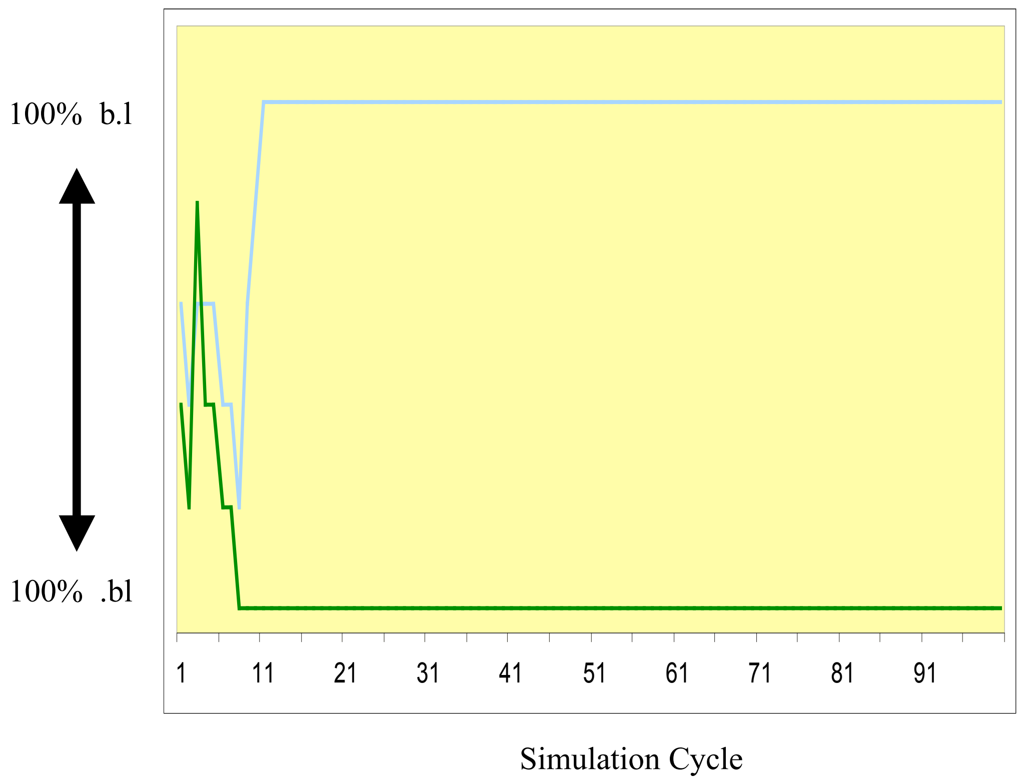
round of the simulation, the set of implications derived from the lexicon for each lexical entry will provide no advantage to one syllabification over the other. However, the fact that there is no advantage in the lexicon to one syllabification or the other will not prevent a syllabification from being chosen. Recall that in the process of assembling a production target from the summed votes from all implications, a feature is randomly chosen from each feature slot. Because that process is random, it is very unlikely that the exemplars stored from this first round will preserve the perfect 50:50 balance between the two syllabifications. The moment there is a numerical advantage of one syllabification pattern over the other in some lexical entry, the system will tend to exaggerate that pattern, spreading it first to those lexical entries most similar to that in which the bias originated, and from there to the entire lexicon. Just as we saw with the differently shaded cells in the cellular automata program illustrated in Figure 3C, once a given syllabification has conquered the entire lexicon, there is no mechanism to resurrect the alternative syllabification, and so the lexicon will continue on indefinitely in that form.

Figure 10. The starting lexicon for simulations shown in Figures 11-13.



Two different runs with this lexicon are shown in Figure 11. The ordinate represents the syllabification of production targets from the two lexical entries labeled *ubli*, where the top of the scale represents 100% [ub.li] targets, and the bottom represents %100 [u.bli] targets. The number of rounds is given on the abscissa. Because in the initial lexicon, all paired lexical entries (including the two *ubli*s) are seeded with exemplars with opposite syllabifications, the outputs of the initial round should cluster around 50% [ub.li], [u.bli], as can be seen to be the case in both runs. However, as suggested above, any departure from 50% within any pair of lexical entries should rapidly push them to jointly settle on one syllabification or the other. All of the lexical entries share at least some features with one another, with the result that most implications derived from the lexicon will apply to many or all of the lexical entries, leading the entire lexicon to eventually veer toward a common syllabification as different syllabifications within lexical entries compete with one another. This can be seen in both runs of the simulation in Figure 11, where a global syllabification for the entire lexicon is rapidly reached, even if it differed from an early trend within the *ubli* pair. (The behavior of other lexical entries is not shown here, as they all reach the same consensus syllabification at approximately the same time.)

Figure 11. Evolution of output forms in the absence of performance bias.

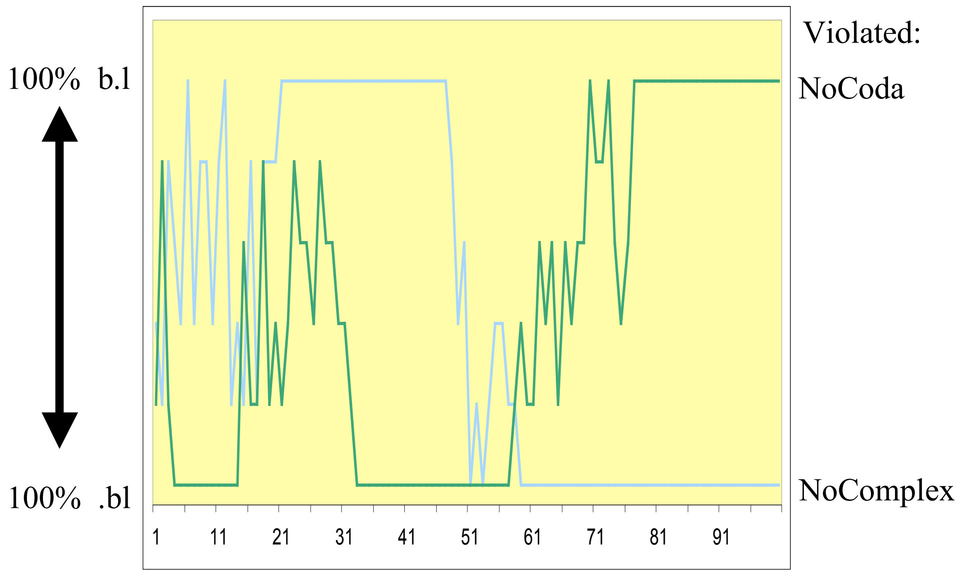


Note that the two runs shown converge on opposite syllabifications. Although there is no way to predict ahead of time which syllabification will sweep the lexicon, because the mechanism of target assembly produces pattern reinforcement, we *can* predict that sooner or later one syllabification will indeed win out. Put in other terms, the two competing syllabification patterns will spontaneously evolve to exhibit a dominance relation. This is precisely analogous to the cellular automata simulation shown in Figure 3, which though begun with equal numbers of light and dark squares, will always end up uniformly one or the other color. This simulation then, was simply a fancier way of showing again that categorical behavior is always a basin of attraction in analogical systems.

1.8.2 Addition of external noise results in oscillation between extremes.

What happens when we supplement the simulation with biases in performance? The following simulation is run with the same lexicon shown in Figure 11, but with addition of two biases in the performance filter, one against codas, and the other against complex onsets. The first changes any word internal coda into an onset, at a rate of 10%, while the other changes any word internal onset followed by a consonant into a coda, again at a rate of 10%. All the possible production targets that this lexicon can produce are therefore going to violate one of these biases or the other, leading to the addition of balanced, but stochastic noise in performance. Results of two runs of this simulation are shown in Figure 12, where again, the lines represent the percentages of the performance targets of the two *ubli* lexical entries that have one or the other syllabification, thereby violating one or the other bias.

Figure 12. Evolution of output forms under performance biases.



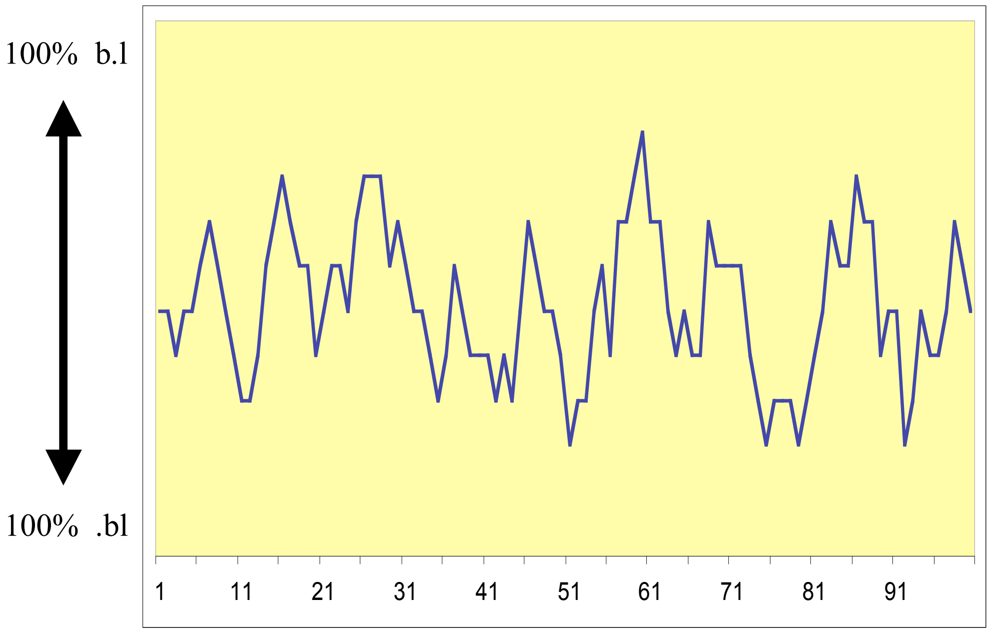
Simulation Cycle

The behavior of the system is similar to that without biases in performance, except that instead of getting locked into one syllabification, the lexicon now oscillates between categorical extremes in syllabification. Restated in terms of dominance, the dominance relation between the two syllabification patterns now alternates between the two possibilities. The ability to emerge from a state in which all exemplars uniformly display one or the other syllabification derives from the biases we added into performance, which periodically alter production targets by changing their syllabification. However, when the lexicon has categorically tilted toward one syllabification, only one of the two biases is active, because every single production target has the same syllabification. Every once in a while then, the active bias succeeds in altering enough of the outputs to allow the other syllabification to gain a foothold in the lexicon again, leading to a possible change in global behavior. Note that because the performance biases are evenly matched, when the lexicon is split between both syllabifications, the biases do nothing more than add extra noise, pushing the system in both directions evenly. The presence of two matched biases that act against each other should then in fact promote a lexicon that tends toward being evenly split between the two syllabifications. The fact that rather than oscillating near the mean, the lexicon oscillates between extreme, categorical behavior reveals that the leveling tendency in the production target assembly mechanism is dominant in this system.

1.8.3 Categorical behavior is dependent on cross-category pattern reinforcement

To show that development of global categorical behavior is in fact crucially dependent on pattern reinforcement in this system, the following simulation was carried out with all ties between and within lexical entries severed: each lexical entry reproduces itself on the basis of one associated exemplar, with no reference to what any other lexical entry has done. The results, shown in Figure 13, show precisely what we predicted in the previous paragraph: in the presence of evenly matched, contradictory biases, syllabification behavior simply oscillates around the mean, while the extremes are avoided. In other words, the dominance relation is dependent on the influence of pattern reinforcement on lexical evolution.

Figure 13. Evolution of output forms in the absence of pattern reinforcement.



Simulation Cycle

Performance Biases: NoCoda = 0.1, NoComplex = 0.1

1.8.4 Interim summary

In sections 1.8.1-3 (as well as in the cellular automata example in Figure 3), we saw that idiosyncrasy is unstable in any system in which an individual element is reproduced under pressure to become more similar to other elements. In such a system, loss of distinctiveness in behavior advances locally, resulting first in regions of identity in behavior separated by sharp boundaries, followed eventually by complete loss of distinctions throughout the system.

The simulation results presented above support the prediction that the domination of one possible phonological pattern over another will constitute a basin of attraction in a system in which production targets are constructed by reference to multiple forms at multiple levels of structure. In such a system, recursion to the mean sets up positive feedback loops which have the following consequences for lexicon structure:

i. Variability within a lexical entry is an inherently unstable state.

ii. Conflicting constraints on output form will tend to exhibit a dominance relation, even when evenly matched in their influence on outputs.

iii. Dominance relations between patterns introduced by closely matched performance biases will oscillate back and forth over time, where periods of categorical dominance are separated by periods of output variability.

1.9 The development of strict pattern domination

Optimality Theory's restrictiveness 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. 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 can often be found paraphrased informally as, 'Lower ranked constraints can't gang up against a higher ranked constraint'. An equivalent formulation that will be especially useful here can be given as:

- The outcome of multiple constraint conflict reproduces the outcomes of the component pairwise constraint conflicts.

These limitations allow OT to predict that certain patterns cannot exist. Just for the purposes of illustration, assume a Markedness constraint banning breathy-voiced vowels (abbreviated NOBREATHY), and another banning mid-round vowels (abbreviated NOMIDROUND). 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 voice and mid-round vowels:

Table 1. Factorial Typology

Factorial Typology: FAITH X NOMIDROUND, NOBREATHY

	NOBREATHY >> FAITH	FAITH >> NOBREATHY
NOMIDROUND >> FAITH	no breathy vowels no mid-round vowels	breathy vowels OK no mid-round vowels
FAITH >> NOMIDROUND	no breathy vowels mid-round vowels OK	breathy vowels OK mid-round vowels OK

This typology shows us that this palette of constraints can be used to describe languages that allow or disallow breathy vowels and mid-round vowels, but cannot be used to describe a language that allows breathy vowels and mid-round vowels in general, but draws the line at vowels that are *both* mid-round and breathy[5]. And in fact, we find that phonological systems do tend to be coarse-grained in this way, drawing fewer distinctions than would seem possible (Pierrehumbert 2002, Gordon 2002a). The principle of strict domination functions to prevent just this sort of interaction between constraints potentially active within a grammar, and its ability to do so arguably underlies much of the descriptive success of OT.

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 in OT are directly or indirectly related to performance biases external to the grammar (e.g., Steriade 1999, Hayes 1999b, Padgett 2002, Hayes and Steriade 2003). This unease arises

[5] If such a language were found, OT could accommodate it by postulating a further constraint against breathy, mid-round vowels, or through the essentially equivalent device of conjoining the two constraints NoBreathy and NoMidRound and ranking the new constraint or the conjunction above Faith

because it is difficult to see how biases with bases 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 difficult than full voicing, and a mid-round vowel is more articulatorily difficult than a high-round vowel, then these costs should compound at some level: a mid-round, breathy vowel should be harder *in toto* than either a high-round breathy vowel or a mid-round voiced vowel. Why then, should increasing difficulty in total articulatory cost through bias-compounding be only sometimes reflected in the grammar, as predicted by the architecture of OT?

The failure of grammars to reflect many of the possible levels of markedness interaction can be restated as a failure of grammatical patterns to reflect the fine-grained distinctions in difficulty that must exist (Pierrehumbert 2002, Gordon 2002a). We saw above that when we model the effect of two opposing biases in an analogically structured lexicon, categorical behavior emerges from gradience. In the following sections, we will see that when multiple interacting constraints are modeled, similar categoricity evolves as well, producing grammatical behavior consonant with the strict domination

---

(e.g., Ito and Mester (2003)). This comes at the cost of essentially adding a principle of OT. In particular, when multiple patterns potentially conflict in a single output form, we will see that categoricity in the lexicon tends to promote outcomes that follow the individual pairwise outcomes of pattern conflict, just as predicted by the principle of strict domination.

1.9.1 Setting up multiple constraint conflict in the simulation

The simulations described in this section are based on a lexicon with three classes of lexical entries, of the shapes VCCV, VC + V and VC + CV, where '+' represents a morpheme boundary. Lexical entries in the latter two classes can be thought of as comprising a stem followed by a suffix. For expository ease, I'll refer to these classes by the class members, 'ubli', 'ip + i' and 'ip + ra' respectively. The lexicon used in the simulations throughout section 1.9 consists of 12 lexical entries in each class, each seeded with five exemplars; the abstract level entries are illustrated below. Each simulation-type shown below was run from the full range of possible seed syllabifications to ensure that all final states shown represent globally accessible minima. The simulations shown in this section were all seeded with 50% of the exemplars showing one possible syllabification, and 50% the other.

---

constraint to the possible set.

Table 2. 'ubli', 'ip + i' and 'ip + ra' lexical entry classes

/ubli/ class	/ip + i/ class	/ip + ra/ class
u b l i	i p + i	i p + r a
u g r a	i d + i	a p + r a
i b l u	a p + i	u g + r a
u p r i	u b + i	i b + r a
u k r a	a p + i	a k + r a
a p l u	u k + i	u b + r a
i g l i	i p + a	i p + l i
u g r u	a d + a	a p + l i
a k r i	u p + a	u g + l i
u k l u	i g + a	i b + l i
a d r i	a b + a	a k + l i
u d l u	u d + a	u b + l i

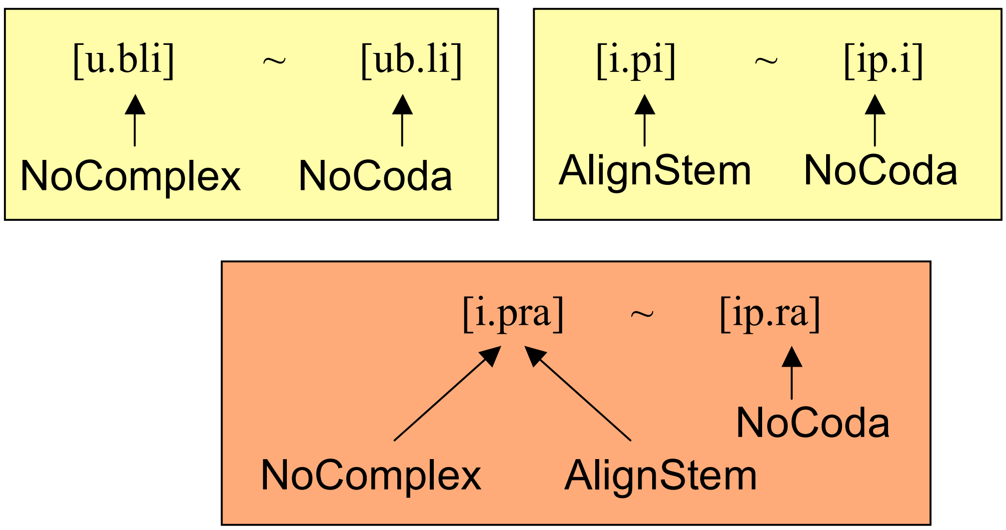
The performance filter is outfitted with three biases, two of which are familiar from the previous simulations, NoCoda and NoComplex. A third bias, abbreviated AlignStem, states that morphological and prosodic boundaries prefer to coincide. This bias operates in performance to shift a syllable boundary to coincide with a stem boundary[6]. The different possible syllabifications for each word-class, and the biases that are triggered by each

---

[6] The choice of 'constraints' to model in these simulations was limited by the fact that this simulation can only model patterns deriving from Markedness, not Faithfulness. The biases chosen here, NoCoda, NoComplex, and AlignStem, were modeled because both pairwise and three-way conflicts can

are shown in Figure 14. In particular, note that for the ‘ip + ra’ class, one possible syllabification triggers both the NoComplex and the AlignStem biases, while the alternative triggers only the NoCoda bias.

Figure 14. Biases triggered by different syllabifications in different word classes



This situation gives us a chance to test the simulation for its ability to reproduce strict domination patterns. Since for the ‘ip + ra’ word class, both NoComplex and AlignStem are triggered by the same syllabification, these

---

be created using the syllabification ‘features’ already in place in the biases both contribute to the total ill-formedness of that syllabification. Hence, these biases should jointly contribute to the pressure to grammaticalize the alternative, NoCoda-violating syllabification.

To give these biases an opportunity to do so, the simulation operates with the bias strengths set such that NoCoda is a stronger bias than either NoComplex or AlignStem alone, but that the latter two together outweigh NoCoda, as illustrated below in Table 3.

---

simulation.

Table 3. Bias Strengths.

Bias strengths	
NoCoda	= 0.1
NoComplex	= 0.07
AlignStem	= 0.07

Expected relative bias in performance			Performance favors:
NoCoda	>	NoComplex	[u.bli] over [ub.li]
NoCoda	>	AlignStem	[i.pi] over [ip.i]
NoCoda	<	{NoComplex + AlignStem}	[ip.ra] over [i.pra]

To show that these relative bias strengths do in fact result in the expected biases in output form after production targets have been filtered through performance, we can look at runs of the simulation with lexicons comprising just isolated word classes.

Figure 15a and b show the results of a typical simulation carried out with a lexicon containing only the /ubli/ and /ip + i/ lexical classes, in the absence of any lexical categories in the /ip + ra/ class[7]. Figure 15a shows simulation results for the /ubli/ and /ip + i/ lexical categories over fifty cycles, while 15b shows the averaged behavior for all twelve lexical categories in each of the two lexical classes. All simulations shown in section 1.9 begin with lexicons seeded with an even mix of the possible syllabifications. Each simulation was done multiple times starting from a range of possible starting syllabifications with similar results, suggesting that the results shown indeed represent universally accessible minima within the simulation space.

[7] Within these and subsequent simulations in section 1.9, the strength of all biases is increased three-fold at the beginning of the simulation, and then is allowed to revert slowly to the default strength over the first 15 cycles. This initial high bias strength serves to create a large degree of variation in syllabification within the exemplars stored in the lexicon, from which patterns can then emerge. Simulations without this initial 'heating' phase eventually produce the same patterns, but can take a much longer time to do so.

Figure 15a. Syllabification of the /ubli/ and /ip + i/ classes:

/ubli/ and /ip + i/.

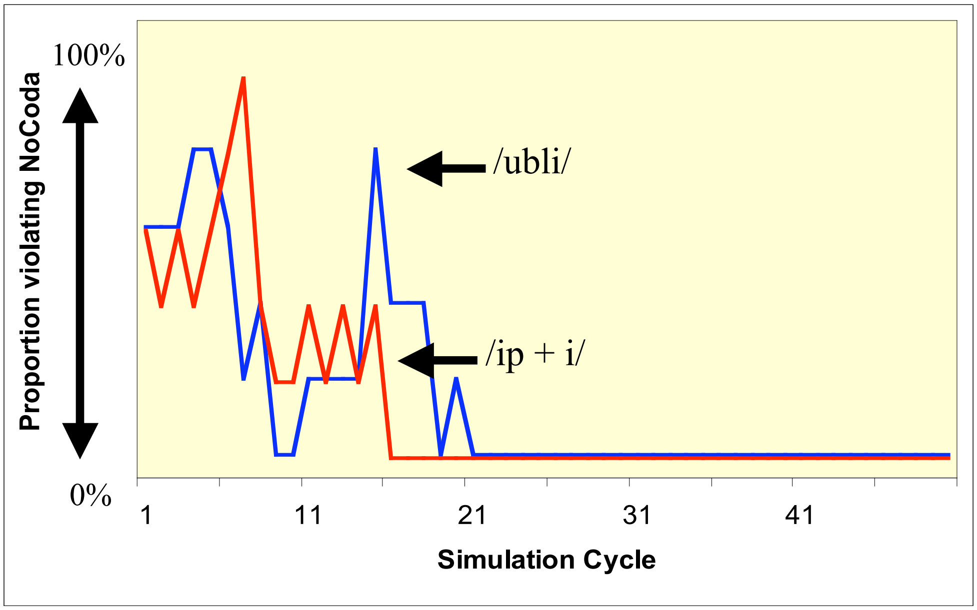
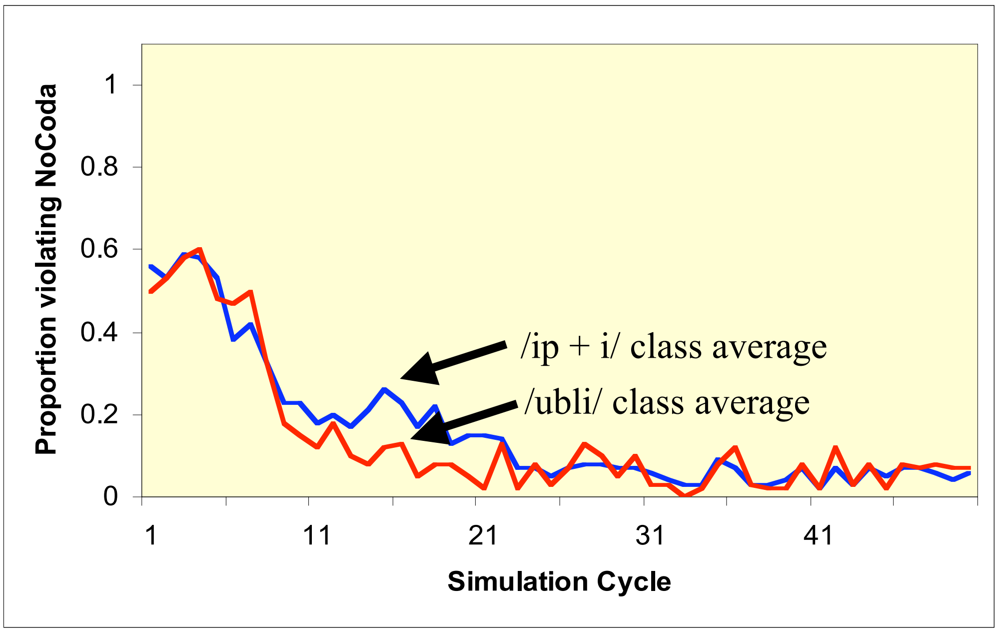


Figure 15b. Syllabification of the /ubli/ and /ip + i/ classes: /ubli/ and /ip + i/ class averages.



- [u.bli] is a stable output. ⇨ NoCoda > NoComplex
- [i.pi] is a stable output. ⇨ NoCoda > AlignStem

In the simulation shown in Figure 15[8], we can see that these word classes rapidly evolve to exhibit syllabifications resulting in complex onsets, and mis-aligned stems rather than codas, as we expect given the greater relative strength of NoCoda in the performance filter, relative to either NoComplex or AlignStem. Recall from the simulations shown above in Figure11 (section 1.8.1), cross-category reinforcement in the assembly of production targets creates a strong tendency to categorical behavior in the lexicon regardless of any bias in performance. However because production targets containing codas are more often altered in performance than those with complex onsets or mis-aligned stems, the performance filter provides a net bias in the direction of coda-less outputs, and thereby makes categorical avoidance of codas in the lexicon more likely than categorical production of forms with codas.

In table 3 above, we made a different prediction above regarding the behavior of the /ip + ra/ class of lexical items, however, because we know that the separate biases against complex onsets and misaligned stems both can

---

[8] The average proportion violating NoCoda never reaches either zero or one, as in Figure 15b, because random changes in performance always ensure that there are some exemplars with the alternate syllabification present in the lexicon.

apply in this class in favor of outputs with codas. If their combined influence result in a coda syllabification in this class more often than NoCoda can produce the alternative, the lexicon's emergent categorical behavior should most often result in assembly of production targets with codas. A typical simulation run with a lexicon comprising solely lexical entries from the /ip + ra/ class is shown in Figure 16a and b below. In accordance with our prediction, we see that the lexicon quickly evolves to incorporate Performance's preference for codas in lexical entries of this class, assembling production targets for all lexical entries with a syllabification that satisfies both NoComplex and AlignStem, while violating NoCoda.

Figure 16a. Syllabification of the /ip + ra/ class: /ip + ra/

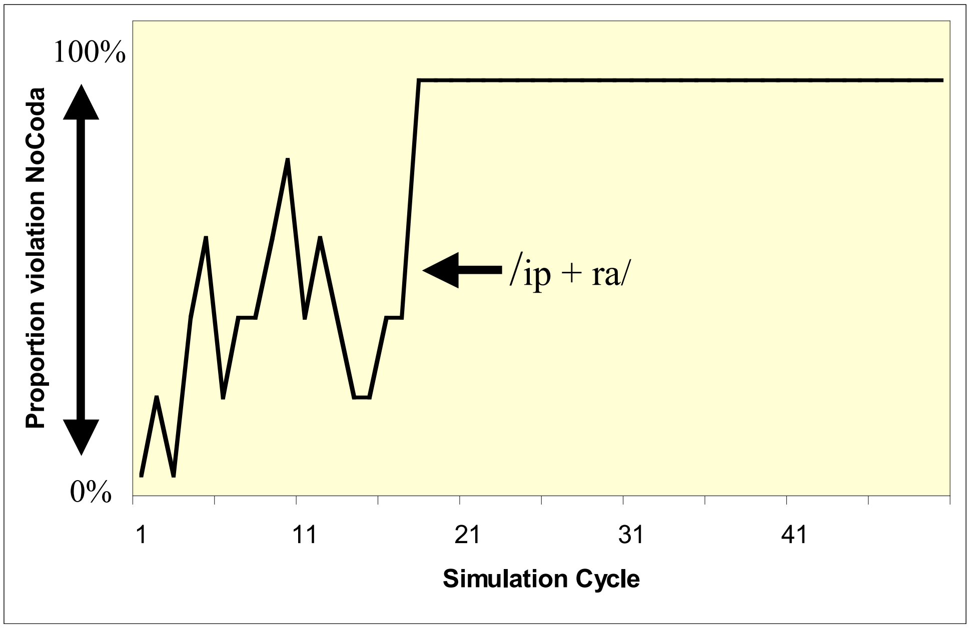
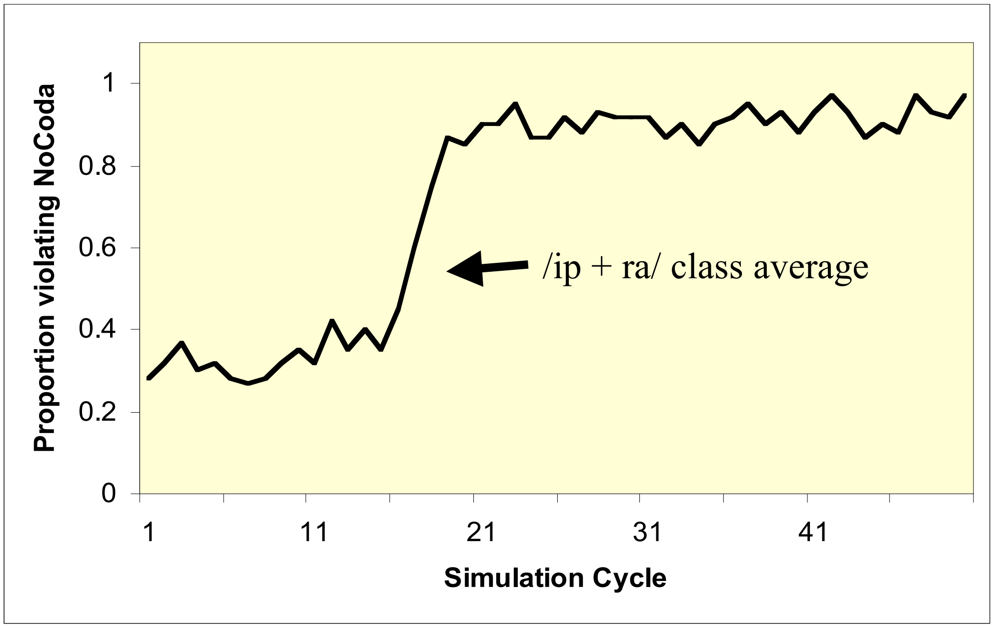


Figure 16b. Syllabification of the /ip + ra/ class: /ip + ra/ class average.



- [ip.ra] is the stable output. ⇒ {NoComplex + AlignStem} > NoCoda

So far then, we see that when the lexicon contains only lexical entries of the /ubli/ and /ip + i/ classes (Figure 15), the NoCoda pattern dominates the NoComplex and AlignStem patterns. On the other hand, when the lexicon contains only lexical entries of the /ip + ra/ class (Figure 14), the joint NoComplex and AlignStem patterns can act together to result in domination of the NoCoda pattern, violating the principle of strict domination. But what happens in a lexicon containing all three classes, where these two possible patterns will come into conflict? A representative simulation with a lexicon containing all three classes together is shown in Figure 15a and b.

Figure 17. Syllabification of the /ubli/, /ip + i/ and /ip + ra/ classes together: /ubli/, /ip + i/ and /ip + ra/.

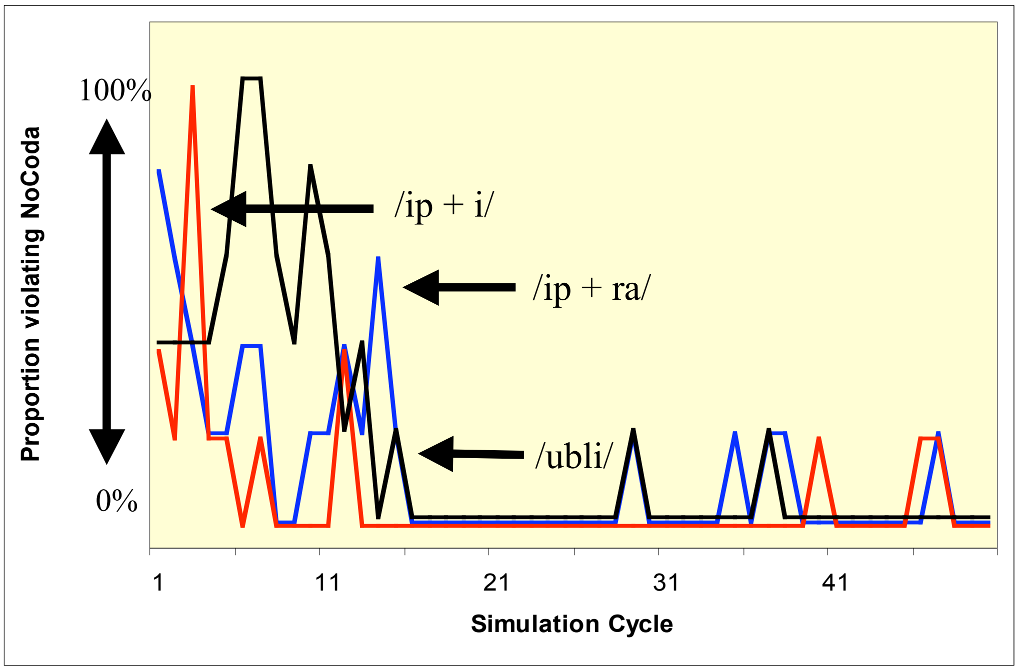
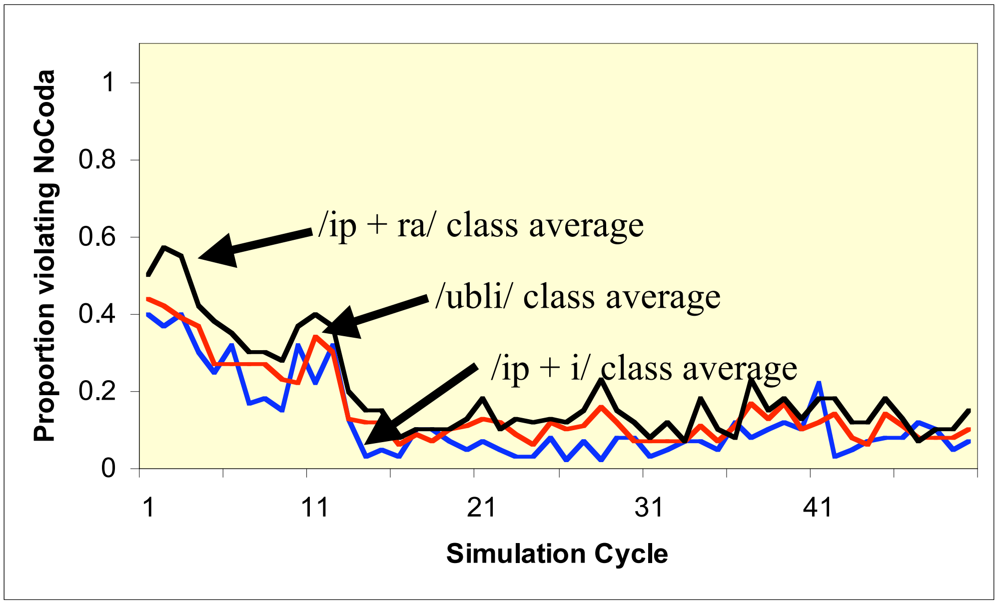


Figure 17. Syllabification of the /ubli/, /ip + i/ and /ip + ra/ classes together: /ubli/, /ip + i/ and /ip + ra/ class averages.



- [u.bli], [i.pa] and [i.pra] are stable outputs ➡ NoCoda > {AlignStem + NoComplex}

As before, the /ubli/ and /ip +i/ classes evolve to categorically assemble production targets without codas. However, we find that now the /ip + ra/ class does likewise, contradicting its behavior when evolving alone. This particular pattern is the most stable state under these particular conditions: in 100 independent runs of the simulation-type shown in Figure 17, at cycle 100, 86 of the lexicons showed the NoCoda > {NoComplex + AlignStem} 'strict-domination' pattern as in Figure 17, while 12 lexicons showed the opposite NoComplex, AlignStem > NoCoda pattern[9].

The finding that there is a minority pattern that is stable within the simulation is not surprising. Because (i) the simulation architecture is non-deterministic, and (ii) categoricity in and of itself contributes to the stability of a state in the system, we expect that the system may randomly explore a number of different possible categorical states, some of which may satisfy performance biases sufficiently to be locally stable. The less frequent, but consistent entrenchment of the NoComplex, AlignStem > NoCoda pattern under these conditions suggests that the conflict between biases against codas, complex onsets and misaligned stems is closely enough matched enough in this lexicon that pressure towards categoricity renders this minority pattern a significant local minimum state. (For more discussion of the influence of the lexicon on the relative stability of this pattern, see section 1.10 below).

Optimality Theory posits strict domination to account for grammatical phenomena in which an outcome of conflict between multiple patterns conforms to the outcomes of the pairwise pattern conflicts. This is parallel to kind of conflict outcome we see in the majority case above. However, there also appear to be grammatical phenomena in which patterns do appear to be able to 'gang-up' to produce an outcome of multiple pattern conflict that appear to derive from additive interactions (see Moreton and Smolensky (2002) for examples), just as seen in the minority case above. To accommodate these latter cases, an alternative to strict domination is provided within Optimality Theory through 'local conjunction' of constraints (Smolensky 1993), in which two independent constraints are conjoined into a single constraint which is violated only if both component constraints are violated within a specified domain. A conjoined constraint can be ranked distinctly from its component constraints, allowing the combined violation of the two component constraints to have an independent effect on optimal

---

[9] Two lexicons were in transition between states, characterized by

output form. To take an example from German discussed in Ito and Mester (2003b), the crosslinguistically attested Markedness constraint banning voiced obstruents, *[+voi. –son] can be conjoined with the constraint banning codas, NOCODA, to arrive at a constraint that is effectively violated only by a voiced obstruent coda. By ranking the simplex Markedness constraints below the Faithfulness constraint IDENT, but the conjoined Markedness constraint above, we account for the fact that German allows voiced obstruents, and allows codas, but does not allow voiced obstruent codas.

Pattern reinforcement across the lexicon in simulations of the model presented here can produce either pattern, depending on the relative strength of biases, the relative typological frequencies of forms affected by those biases (see section 1.10 below for further discussion of this point), and also depending simply on the idiosyncratic history of the system. As we saw above, out of 100 independent runs of the same simulation, starting from the same point each time, two distinct stable states could be reached, only one of which would be described in OT through the mechanism of 'strict domination'. By changing the bias strengths or the relative numbers of lexical items, it would also be in principle possible to see lexicons emerge in which

---

idiosyncratic behavior across lexical categories.

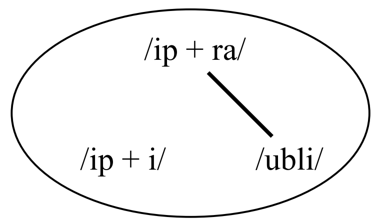
NoCoda dominates *Complex and AlignStem independently, producing stable outputs [u.bli] and [i.pi], but where *Complex and AlignStem together dominate NoCoda, producing the stable output [ip.ra]. This state of affairs would be described in OT through the mechanism of constraint conjunction. The virtue of the model presented here is not that it can account for lexicons in which biases contribute additively to output forms, but rather that it can also account for the evolution of lexicons in which biases do not appear to contribute additively, i.e., in a manner predicted by the OT principle of strict domination. The property of the model that allows this failure of additive behavior to develop lies in leveling through pattern reinforcement within the lexicon.

To see this better, let's return to the majority pattern described above, where we find that the NoCoda pattern is dominant in the simulations with all three lexical classes present. The crucial difference between the simulation in Figure 16 and 17 lies in the fact that in the latter, the /ip + ra/ class evolves not influenced only by the net bias in performance for output forms with codas in its own class, but also influenced indirectly by performance's bias for forms *without* codas in the other two classes. In this model, both the preferences for and against the NoCoda pattern in different classes arise in the performance filter, but the fact that members of the three lexical classes share features with one another allows behavior in one class to influence behavior in another. The pathways in this lexicon through which performance biases can influence patterns are illustrated below in Figure 18.

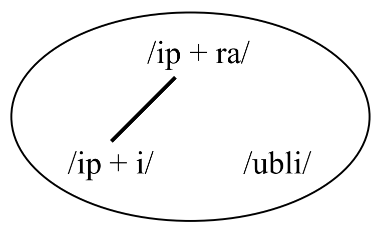
Figure 18. Pathways of Bias influence.

Lexical classes that are influenced by the bias shown are connected by a solid bar.

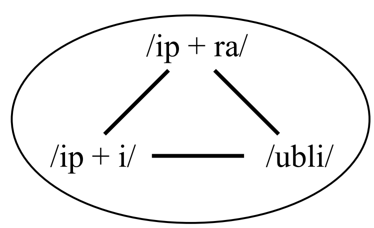
a) NoComplex:



b) AlignStem:



c) NoCoda:



The important point to take from Figure 18 is that there are more opportunities in our lexicon for the bias against codas to influence syllabification than there are for the other two biases to influence syllabification. Because every lexical entry in this lexicon shares a feature set serving as a context that can be affected by NoCoda, every time performance returns an output satisfying the bias against codas in one form, it predisposes not only that lexical entry to reproduce that bias, but every other entry in this lexicon as well. In contrast, when an output is returned that satisfies NoComplex or AlignStem, it affects a smaller portion of the lexicon.

This result can be restated in the following way: because *all* entries in this lexicon share the feature sequence that results in avoidance of codas in the outputs of *some* lexical entries, *all* entries are indirectly under pressure from performance to reproduce outputs without codas, through the fact that old output forms influence the pattern of new output forms. Thus, although the additive effect of the performance biases against complex onsets and mis-aligned stems consistently biases tokens of the /ip-ra/ class in favor of a syllabification with a coda (e.g., [ip.ra]), the influence of the lexical entries in the other two classes will bias the lexicon toward outputs with a shape [i.pra]. If the influence of pattern reinforcement is stronger than that of performance, as in this simulation, the result is a lexicon that tends to produce outputs in accordance with strict domination, in spite of the fact that the biases in performance do interact additively at the level of performance. In this model then, strict domination is not a fundamental property of the grammar, but is a derived result of the mutual influence of lexical entries on one another.

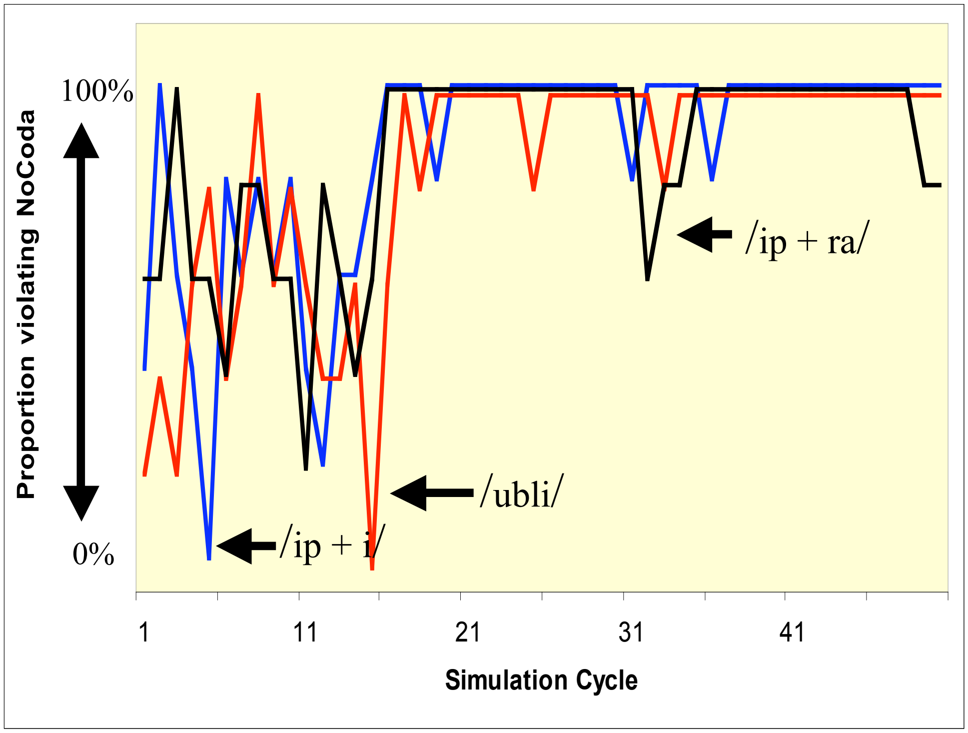
Note that the lexicon can self-organize in this way only when pattern reinforcement between lexical items is strong enough to have a significant effect on output form relative to that exerted by biases in performance. If the influence of conflicting biases is increased in these simulations beyond a given point, the outputs for each lexical entry are essentially randomized in each round, preventing structure from building up to influence future cycles. If you will recall, one of the criteria for self-organization is that the system must exhibit hysteresis, that is, it must fail to reach equilibrium between cycles of structure formation. In this simulation, this can only be achieved when a given bias in performance is weak enough that it fails to modify most sequences that contravene that bias in a given round.

1.10 Type frequency and pattern dominance

As we saw in the last section, the effect of performance biases on the production targets of each lexical entry have a cumulative influence on the evolution of the entire lexicon. Hence, type frequency should have a significant influence on the evolution of grammatical patterns, leading to the prediction that the larger the number of lexical entries a bias potentially applies to, the more likely its effects will be reflected in the behavior of the lexicon as a whole.

To test this hypothesis within the simulation, we can increase the number of lexical entries in the /ip + ra/ class relative to the other two classes, to allow the combined performance biases against complex onsets and misaligned stems more opportunities for influence. In the simulation shown in Figure 19, the number of lexical entries in the /ip + ra/ class was increased to 20, while the number of lexical entries in the /ubli/ and /ip + i/ classes were decreased to 5 each. In this lexicon then, there are twice as many lexical entries that can be influenced by both NoComplex and AlignStem together, as there are lexical entries that can be only influenced by one of these biases.

Figure 19. Syllabification of /ubli, /ip + i/ and /ip + ra/ classes with a 4x higher /ip + ra/ relative type frequency.



- [ub.li], [ip.i], [ip.ra] are stable outputs. ⇨

  AlignStem, NoComplex > NoCoda

In Figure 19, we see that in contrast to the simulation run on the typologically more balanced lexicon shown in Figure 18 above, the lexicon as a whole rapidly evolves to assemble production targets with codas, to the exclusion of complex onset and mis-aligned stems. An OT analysis of this pattern using the three 'constraints' that we've placed into the performance filter would find that NoComplex and AlignStem both independently dominate NoCoda – but we know that in this simulation the bias against codas is actually stronger than either the bias against complex onsets or misaligned stems. Here, it is the fact that these two biases can act together in a large proportion of the lexicon that allows them to jointly determine the behavior of the entire lexicon. This example illustrates the fact that in this model, the relative 'strength' of biases in performance may have less effect on the propagation of patterns through the lexicon than the relative distribution of feature sequence types within the lexicon itself.

1.11 Alternative mechanisms for assessing similarity.

The development of categorical behavior across lexical categories in this system derives from cross-category influence in the assembly of production targets, in conjunction with re-storage of outputs back in the category of origin. Similarity-based cross-category influence results in 'blending inheritance' in output forms, which must result in steady erosion of difference across the lexicon, and re-storage of forms creates a feedback loop, which together result in increased stability of categorical behavior over time relative to non-categorical behavior (Cooper 1999, Abler 1989). As long as these conditions hold, categoricity should be more stable than variation over time. Alternative mechanisms for assessing similarity, then, should still produce qualitatively similar results, differing only, for example, in their transition breakpoints between alternative patterns of conflict resolution.

In the simulations shown in sections 1.9 and 1.10 above, the implication set derived from patterns in the lexicon was used without further modification to assess similarity between lexical categories. Because the implication set records patterns in the lexicon in terms of their frequency of occurrence, direct application of the implication set without any further modification results in influence of target assembly by patterns in the lexicon in direct proportion to their type frequency. Frequency of a generalization is certainly not the only way to assess similarity, however. To explore the ramifications of using different mechanisms for determining similarity, additional algorithms based in (i) generalization reliability (Albright 2002, Albright and Hayes 2002), and (ii) Skousen's Analogical Language Modeling algorithm (Skousen 1989, 1992) were implemented and investigated.

1.11.1 Frequency

To provide a basis for comparison, this section briefly recaps the mechanism for assessing frequency based directly in generalization frequency. In the simulations shown in sections 1.9-10 above, generalizations contributed to output forms in direct proportion to their frequency. When implications are applied to target assembly on the basis of frequency, no assessment is made concerning how informative that implication is. For example, an implication of the form, [C] → [C, onset] may be very frequent in the lexicon, but if the implication [C] → [C, coda] is just as well represented, then neither of these implications has any predictive value in telling us whether a given underlying consonant should surface as an onset or coda. In the frequency-driven version of the program, both of these implications would contribute equally to influencing the possible behavior of an underlying consonant, with the result that the these implications provide no net bias on the syllabification of a consonant. (Note that in this simulation, no allowance is made for the possibility that distinct features competing for the same slot might be able to interact – rather, they simply cancel each other out.) The program running the simulation is equipped to apply the implication set to target assembly in proportion to any exponential scaling of frequency. In the examples above, frequency was linearly scaled, such that influence of frequency was directly proportional. Small deviations from direct proportionality were also tested, with no qualitative difference in the character of results (not shown).

1.11.2 Reliability

Alternatively, the tendency of conflicting implications to cancel one another out in target assembly can be modeled by calculating the *reliability* of an implication, and using this reliability measure as the factor determining the relative influence of each implication on target assembly. For example, if the antecedent for an underlying feature sequence is shared by multiple consequents, meaning that there is no clear preference in the lexicon for a single mapping between a given underlying feature sequence and an output, then we can say that there is no *reliable* generalization over the lexicon concerning mappings from that underlying feature sequence to outputs (for a discussion of reliability of generalization, see Albright 2002, Albright and Hayes, 2002). If, on the other hand, there is only one consequent associated with a given antecedent, the implication represents a highly reliable generalization[10]. For example, if an antecedent is associated with only one consequent, that implication (the antecedent-consequent pair) is assigned a reliability of 1. If an antecedent is associated with two consequents, each antecedent-consequent pair is assigned a reliability of 0.5.

Reliability factors may be further raised to a specified exponent to allow non-linear scaling of influence in target assembly by reliability factor. If an exponent of 1 is used, for example, an implication with a reliability of 1 (i.e., one that is completely reliable) would have twice the influence on target assembly as an implication with a reliability of 0.5. Larger exponents increase the influence of more reliable implications relative to less reliable implications; very large factors essentially limit influence in target assembly to completely reliable implications.

In the context of these simulations, the assessment of similarity by frequency versus reliability differs primarily in that in the former, type frequency of a pattern (i.e., the number of separate lexical items exhibiting the

---

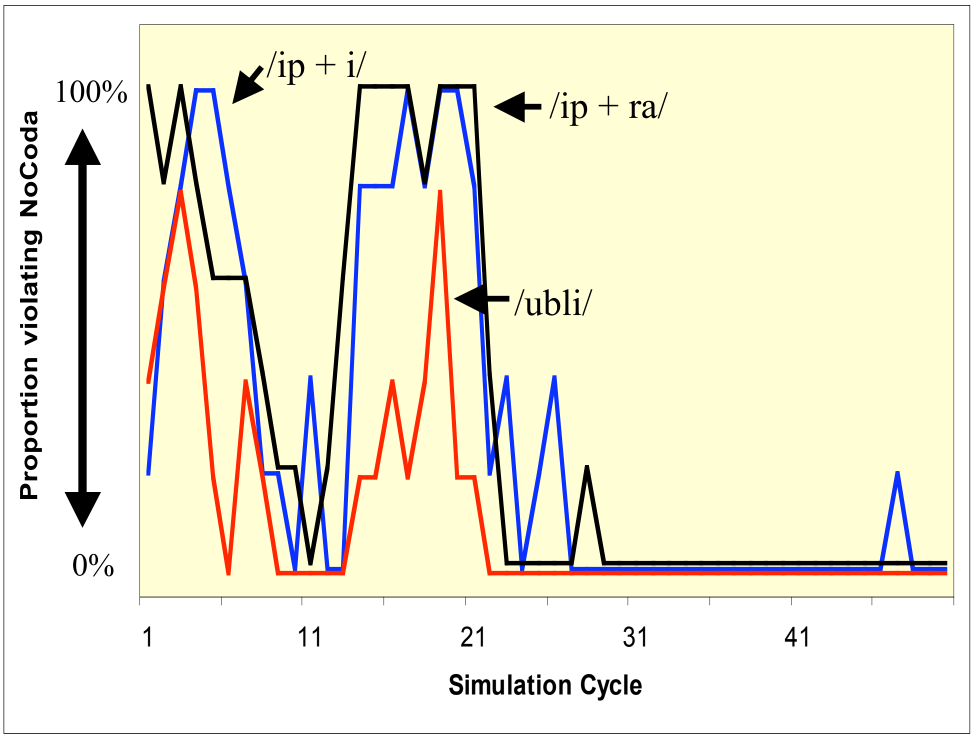
[10] If an antecedent-consequent pair occurs only once in the lexicon, then it will be of necessity 100% reliable. Other work has included a factor of type frequency in the calculation of 'reliability' to limit the influence of these incidentally reliable, low frequency generalizations (see Albright 2002,

pattern) is a direct factor in the strength of a pattern. In the latter, type frequency is not directly measured; instead only the degree of adherence to a given pattern is measured, that is, the relevant factor is the degree of pattern consistency within the set of lexical entries that *could* show the pattern, irrespective how large that set is.

Figure 20 below shows results from a reliability-based simulation run with the same starting conditions as that in Figure 17 above, in which lexical entries of the form /ubli/ /ip + a/ and /ip + ra/ all evolve in tandem, under the influence of NoComplex (bias strength = .07), AlignR (bias strength = .07) and NoCoda (bias strength = .10). In the simulation shown, the scaling exponent on reliability strength was 2. Simulations done with subsets of the lexical classes (i.e., /ubli/ and /ip + i/ alone, /ip + ra/ alone) provided qualitatively similar results to those shown in Figures 13-14 above, indicating that changing the mechanism of similarity assessment did not alter the apparent relative strengths of the three performance biases when allowed to interact in isolation (not shown).

Albright and Hayes 2002); here, generalizations that were 100% reliable because they held over only one item were excluded.

Figure 20. Reliability: Syllabification of the /ubli/, /ip + i/ and /ip + ra/ classes together.



1.11.3 Analogical Language Modeling

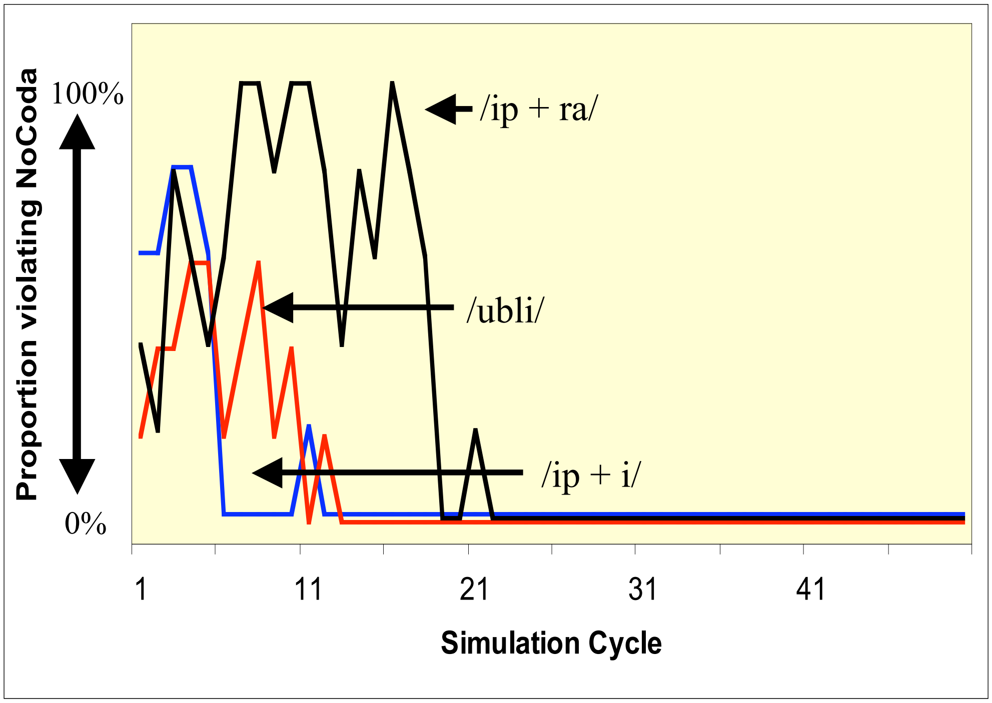
The Analogical Language Modeling (ALM) algorithm developed by Royal Skousen (1989, 1992) has been used extensively to model pattern extension in morphological and morphophonological processes (see, for example, Eddington 2000, 2002, Krott et al. 2001). The effect of the ALM algorithm is to exclude an antecedent A from influencing a particular outcome, if there exists a more specific antecedent B that contains antecedent A, where antecedent B is associated with multiple consequents. Within this simulation, the ALM algorithm was implemented in the following way: for each antecedent in the grammar, identify all other antecedents that *contain* the current antecedent. This set of larger antecedents is the current antecedent's *supracontext*. If each antecedent in the supracontext is associated with one and only one consequent, then the supracontext is *homogeneous*. If the supracontext is homogeneous, the number of times each of the consequents of the current, supracontext-defining antecedent is represented in the supracontext is counted[11], and then each of these consequents is employed as before in target assembly. If the supracontext defined by a given antecedent is not homogeneous, the antecedents within that supracontext do not contribute

---

[11] In ALM, the likelihood of choosing a given feature from a given implication is proportional to the number of *pointers* for that supracontext-defining antecedent (Skousen, 1989, 1992, Eddington 2000); here pointers are calculated as the product of the number of times the antecedent was represented in the lexicon, and the number of members of the supracontext defined by that antecedent. Simulations were also done without conversion of frequency to pointers with indistinguishable results.

to the outcome at that point. Note that every antecedent is likely to be part of many supracontexts; for each supracontext that is homogeneous, each contained antecedent will have an additional opportunity to contribute to the current output form. As a result, depending on the particular patterns extant in the lexicon, a given antecedent may contribute many or few times to an output. Figure 21 shows a typical simulation run, using the same conditions as in the figure above. Again, simulations done with subsets of the lexical classes provided qualitatively similar results to those shown in Figures 15-16 above.

Figure 21. ALM: Syllabification of the /ubli/, /ip + i/ and /ip + ra/ classes together.



Comparison of Figures 19, 20 and 21 shows that the outcome of pattern conflict under all three mechanisms of similarity assessment is qualitatively similar: the coda-less pattern comes to dominate the simulated lexicon, even though the performance bias against codas is not as strong as those against complex onsets and misaligned stems, taken together[12]. The ability to find conditions under which all three mechanisms of similarity assessment result in the same outcomes is expected, because increased stability of categorical solutions to pattern conflict is a necessary result of similarity-based cross-category influence.

The three mechanisms of similarity assessment do result in some quantitatively different behaviors in the simulations, however. For example, within reliability-based simulations, use of scaling exponents between approximately one and four result in increased relative stability of any categorical patterns, once reached. As another example, the stability of any given categorical pattern under the ALM algorithm is considerably more sensitive to the precise relative values of each performance bias, and to the featural makeup of lexical categories, relative to either of the other two methods of similarity assessment. A detailed exploration of the behaviors of each of these similarity assessment mechanisms is beyond the scope of this dissertation, and so no further investigation was undertaken.

---

[12] The lexicon with higher /ip + ra/ class type frequency was also tested with each of these alternative similarity assessment mechanisms, with qualitatively similar results as found in Figure 17 (not shown).

1.12 Conclusions

In a model in which pattern reinforcement across categories in the lexicon and performance bias on the form of individual outputs are able to dynamically interact through feedback, categorical dominance, both relative to pairs of biases, and to larger interacting bias sets, is a spontaneously reached, stable evolutionary state. Neither pattern domination nor strict pattern domination need to be stipulated in the model, as they are straightforward consequences of the structure of similarity relationships within the lexicon in conjunction with a tendency to pattern reinforcement in production. Crucially, the model provides no mechanism that can produce categorical dominance relations between patterns in one step. Rather, the development of categoricity requires feedback, which in turn requires that the system proceed through multiple cycles. The dynamic dependence on the interaction between system elements through feedback is a characteristic of self-organization-driven pattern formation.

Self-organization also requires that elements within the system be able to interact distinctly, as it is the differences in interaction that provide the non-linearities driving pattern formation. In this model, the difference in interaction between system elements is based in degree of similarity.

Consequently, any mechanism for assessing and deploying similarity distinctions between forms should in principle be able to support categorical pattern formation, as found here in the comparison of similarity generalizations stated in terms of frequency, reliability or satisfaction of the constraints of ALM. Therefore, the absence of qualitative differences in these approaches at this level of abstraction cannot be used to make any prediction concerning the likelihood that any one of them will or will not turn out to provide a good description of real-world natural language processing.

Likewise, the success of the model underlying the simulations presented here in producing categorical dominance patterns does not prove that categoricity in phonological systems actually arises in this way. Rather, these results should be taken as an existence proof that some patterns in categoricity of the sort found in phonological systems *can* arise through self-organization within a very general, psycholinguistically plausible model of cyclic production and storage of forms.

The primary thrust of the simulations presented here has concerned the patterns deriving from interactions between lexicon-external influences on lexical form – dubbed performance biases here – and cross-category influence within the output assembly mechanism. However, as it stands the model makes several additional interesting predictions about lexicon-internal influences on the development of categorical patterns.

1.12.1 The influence of type frequency in pattern evolution

As seen in section 1.10 and Figure 19 above, the model predicts that the probability that a bias will eventually result in a categorical pattern is correlated with the frequency of lexical items containing sequences influenced by that bias (e.g., Bybee 2001). The hypothesis that relative type frequency in the lexicon may play a role in grammatical patterning is indirectly supported by Gordon's finding that whether or not CVC syllables count as 'heavy' in a quantity-sensitive stress system depends strongly on the relative proportion of highly sonorant to less sonorant coda segments (Gordon 2002b). In a survey of 62 quantity-sensitive languages differing in their treatment of coda consonants, Gordon found that the languages within the survey were overwhelmingly more likely to treat coda consonants as contributing to syllable weight if higher-sonority codas outnumbered lower sonority codas within the coda inventory, and vice versa. If the size of a natural class within the coda inventory correlates with the type frequency of that class across lexical item, this finding is consistent with the notion that the more opportunities a performance bias has to influence the lexicon, the more likely it is that the grammatical system will evolve to accommodate that bias.

1.12.2 The influence of bias specificity on pattern evolution

If the type frequency of a sequence related to a particular bias correlates with the grammaticalization of that bias within the lexicon, then all else being equal, a relatively general bias should be grammaticalized more easily than a bias referring to a subset of the sequences referred to by the general bias. This prediction derives straightforwardly from the set/subset relation. Feature sets fitting the description […abc…] (where '…' stands for any additional features) can never be **more** frequent than sets fitting the descriptions […ab…] or […bc…], because the feature string *abc* contains the substrings *ab* and *bc*. In fact, sets described by […abc…] can only be equivalently frequent to sets described by […ab…] or […bc…] in the special case in which feature strings *ab* and *bc* never occur outside of the feature string *abc*. In the more general case where substrings of a feature string occur independently of larger strings containing them, sets described by a given feature set will necessarily be fewer in number than the sets described by any of their feature subsets. *Ceteris paribus* then, the more context-sensitive the formulation of a bias, the less likely that particular formulation of the bias is to have a significant influence on the evolution of the lexicon.

It has been frequently remarked that phonological systems make fewer categorical distinctions than might be warranted by the phonetic facts (see, for example, Gordon 2002a for discussion of this issue in terms of the ‘inefficiency’ of phonology in regulating syllable weight distinctions). Pierrehumbert (2002, 2003) provides arguments that the ‘coarseness’ of phonological distinctions 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 an additional mechanism to limit the number and specificity of grammatical patterns. Under the influence of similarity-driven pattern reinforcement in the lexicon, patterns holding over many forms will always tend to absorb conflicting patterns holding over fewer forms (as we saw above in the cellular automata simulation in Figure 3). This constant drift toward pattern consolidation mitigates the tendency of highly context sensitive biases in performance to produce ever greater distinctions in lexical form. The conflict between the two should result in a shifting compromise in which the phonology supports fewer distinctions than would appear optimal given the phonetics.

1.12.3 Some further issues.

Section 1.11.1 above discusses the prediction that the more opportunities a performance bias has to influence the lexicon, the more likely it is that the grammatical system will evolve to accommodate that bias. But taken to a logical extreme, doesn't this seem to suggest, for example, that the more codas a language has, the more likely it is to categorically lose its codas through the influence of a bias against them? Two factors mitigate this prediction in interesting ways. First, the simulations in this chapter treat the values of external performance biases as constants, rather than allowing for the possibility for feedback interactions between performance biases and the lexicon. However, there is no reason that performance biases should be independent of the fact of performance itself. For example, there is every reason to think that the strength of an articulatory bias against a given element will be negatively correlated to some degree with the frequency of that element in production, because practice is expected to render intrinsically hard gestures and gestural combinations easier (Sternberg et al. 1980, Saltzman and

Munhall 1989, Shadmehr and Bashers-Krug 1999, Ussishkin and Wedel 2003 and references therein, Wedel 2004). Because of the inverse relationship between frequency of use and difficulty, this model then predicts that crosslinguistically marked segments or structures will tend to be stable in a language as long as they remain relatively frequent within the experience of the language user. However, if such marked segments or structures become relatively rare within a lexicon through language change, they should become less stable.

Secondly, performance biases against particular output forms are not the only forces influencing lexical patterns. In the parlance of Optimality Theory, constraints against properties of output forms – Markedness constraints – must be balanced by Faithfulness constraints that act to preserve contrast. In the simulations shown in Chapter 1, there is no structure that could play the role of Faithfulness in preserving contrast, suggesting that a major part of the picture is still missing. It will come as no surprise then, that extension of the simulation architecture to model of the role of contrast in shaping lexical patterns forms the subject of the next chapter.

Chapter 2. Categoricity versus contrast in phonological systems

2.1 Introduction

One of the most salient characteristics of linguistic systems are their high degree of categorical behavior, and yet, no languages seem to veer too far towards across-the-board pattern uniformity. From the point of view of the general framework presented in the last chapter, the preservation of many distinct categories is in fact unexpected. Recall from section 1.6.2 in Chapter 1 that systems evolving solely through blending inheritance of any sort tend inevitably toward uniformity, because outputs always lie within the extremes represented by the current state. Each generation of the system therefore risks losing the representation of some extreme value in its current outputs, which then narrows the possible range of variation of all future generations. It is true, of course, that in systems where some interactions are of greater consequence than others (e.g., when local interactions are stronger than non-local interactions), the trajectory from complete disorder to total uniformity is marked by stages in which elements with different behaviors become segregated into sharply bounded regions of similarity (as we saw in section 1.6.2 above). However, in the absence of intervention to introduce or maintain distinctions, patches of uniformity will always tend to either grow at the expense of their neighbors, or be swallowed themselves as the system evolves (see Figures 3C and 11 from Chapter 1). Consequently, even though a system may pass through stages in which many distinct categories co-exist, these are just way-stations on the path to maximum categoricity, that is, uniformity. If we assume, as argued in the first chapter, that analogical pressures (whether from cross-category influence on output form or some other mechanism) influence the evolution of the lexicon, then we seem to be predicting that the end state of lexicon evolution should be a lexicon with a single word-form populating all categories. Hence, the fact that the number of distinct forms distributed across categories does not seem to inexorably decrease over time, suggests that other pressure(s) must exist to preserve categorial distinctions and/or create new ones as previously existing distinctions are lost.

In Chapter 1 above, we identified one mechanism to maintain some local distinctions along the pathway to uniformity, in the form of context-sensitive performance biases. In section 1.8 above, we explored the evolution of a single parameter, the syllable affiliation of consonants in a simulated lexicon in which outputs were derived from corresponding lexical entries under cross-category pressure from other phonologically similar entries. As expected, we found that in the absence of any introduced difference from outside, syllable-affiliation evolved to be uniform. However, we saw that the system could be prevented from reaching a state of unchanging uniformity in syllabification through the occasional imposition of different syllable affiliation values from the outside in the form of performance biases. We further saw that when these biases on output form were made to be context sensitive, stable regions of difference could develop, in the form of particular syllabifications matched to particular contexts.

This is not enough, however, to explain why lexicons never seem to lose all contrast, for the development and preservation of categorical distinctions through context-sensitive performance biases is dependent itself on the existence of a prior distinction in context[13]. At the moment a system loses a difference in context, the ability of performance biases sensitive to that

---

[13] Within the simulations in Chapter 1, the CC sequences that provided the context for the conflict between NoCoda and NoComplex could not themselves be lost. If deletion had been possible, a deletion turning a CC sequence into C would have produced a more stable outcome due to the resulting lack of pressure to alternate from biases in Performance.

context to preserve difference is also eliminated, meaning that performance biases can only slow the loss of categorial contrast, rather than themselves solve the problem of the persistence of pattern multiplicity. In response, this chapter will present results of simulations showing that when categorization is included as a step in the simulation, feedback between production and categorization produces indirect selection for contrastive variants, resulting in preservation of contrast despite the steady tendency of forms to become more similar at the output level.

In addition, a closer look at patterns that emerge in simulations through the interaction of categorization with (i) analogical pressure through cross-category influence on the one hand, and (ii) performance biases on the other, suggest an account for the observation that markedness within and across languages is correlated, that is, that crosslinguistically rare sounds/structures also tend to be relatively rare in the languages that do have them (e.g., Greenberg 1966, Frisch 1996, Kochetov 2002: Chapter 5). For example, contrastive nasality in vowels is crosslinguistically marked (Ferguson 1963). At the same time, if a language does allow a contrast in vowel nasality, nasal vowels are likely to be less common than oral vowels (Ferguson 1963).

Optimality Theory accounts for crosslinguistic patterns in the structures categorically admitted by a grammar, through codifying cross-linguistic tendencies into elements of the phonological system itself in the form of markedness constraints. However, if faithfulness to a marked structure outranks the markedness constraint against it, an Optimality Theoretic grammar provides no straightforward mechanism to discourage that structure from appearing in an underlying form. OT therefore leaves the correlation of crosslinguistic markedness with intra-language markedness unexplained. In addition, there is abundant evidence that phonological systems can include patterns that seem language-specific, as well as patterns that apparently contravene crosslinguistic markedness tendencies (Bach and Harms 1972, Hellberg 1978, Anderson 1981, 1985: pp 78-9, Hayes 1999a). A goal of this chapter will be to show that within the present framework, the feedback-driven link between grammar-external performance biases and pattern reinforcement in the lexicon can account for the statistical link between cross-linguistic and within-language frequency patterns, along with the possibility of exception.

The remainder of the chapter is organized in the following way. Section 2.2 begins with an overview of the role ascribed to contrast in lexical systems and selected theoretical responses, followed by arguments that

Darwinian selection can provide useful insights into maintenance of contrast in lexical systems. Section 2.3 introduces a simulation architecture that includes categorization as a step in the production – storage cycle. The simulation is described conceptually in section 2.3.1, and then described in greater detail in sections 2.3.2-8. In sections 2.4-6, I present results from simulations illustrating that cross-category pattern reinforcement, selection for contrast, and performance biases interact to produce patterns parallel to those discussed above. Finally, section 2.7 discusses these results in the larger context of modeling phonological competence.

2.2 Categorization and contrast

Throughout the history of the study of language, linguists have articulated the intuitively appealing idea that certain language structures are related in one way or another to certain language functions. At a deep enough time-depth, this is trivially true, because those genetically based structures that underlie language (whether language-specific or not) are very likely have emerged through the Darwinian process of natural selection and differential inheritance rather than entirely through accidental spread of random mutation[14]. However, even at the much shallower time-depths of diachronic change and individual language acquisition and use, factors emerging from language use have long been argued to influence and constrain the space in which languages are free to evolve (for a sampling of viewpoints, see Boudoin de Courtenay 1895 [1972], Saussure (in Anderson 1985, p. 40), Ohala 1981, Bybee 1985, Lindblom 1992, Kirby 1999, Hurford 2000, Bybee 2001, Blevins 2003, Mielke 2003).

One such function that has been proposed to influence language form is the requirement for contrast. The communication of information, writ large, is arguably a primary function of language, and information transmission requires detectable contrast between differently signifying message elements

---

[14] In biological evolution, only genotypes that directly or indirectly result in some function (more precisely, some phenotype) are subject to the non-random differential inheritance that results in their enrichment and entrenchment in a population. So-called 'silent' genotypes – those that have no effect on phenotype in a given context – are neutral with regard to evolutionary change in that context. Such silent genotypes can in fact become entrenched in a population through the statistical vagaries of reproduction, as opposed to through phenotypic selection, but this pathway to entrenchment is

(Saussure 1916 (1959), Shannon 1949). A speaker creates a message consisting of encoded categories, and the hearer successfully decodes the message in relation to the degree that s/he can detect those categories as encoded. Because errors in categorization grow more likely the less contrast there is between distinct message elements, reliable message transmission requires a degree of contrast between differently signifying message elements sufficient to allow correct categorization most of the time. The fact that language is regularly used with some degree of success in information transmission therefore indicates that a sufficient degree of contrast is available to distinguish between most distinctly signifying utterances[15].

---

statistically extremely unlikely for a genotype as complex as that underlying language is likely to be.

[15] Does this observation alone indicate that there must be a present causal link between the function of contrast in information transmission and present language form? Not by itself, for if the vocabulary and grammar for all humans were all the same, and we had evidence that they had remained so since the evolution of human language itself, it would be sensible to propose that linguistic contrast, in its current evolved state, no longer retained any causal connection to the function of contrast in communication. However, the fact that vocabularies and grammars are not stable, but rather in constant flux over historical time, suggests that some mechanism(s) must be available to reintroduce contrast into a system to compensate for the contrasts that will

Optimality Theory is structured to address at least part of the requirement for contrast maintenance in language. Within OT, the two originary constraint classes, Faithfulness and Markedness constraints, operate in ranked opposition to one another to produce the range of phonological patterns possible in language. While Markedness constraints are solely sensitive to output form, Faithfulness constraints operate to keep input-output mappings identical, and have been widely described as functioning to preserve contrasts between distinct output forms in synchronic grammars (Prince and Smolensky 1993, Kirchner 1997). However , because Faithfulness constraints are not in fact sensitive to contrast itself, the continual tension between Faithfulness constraints maintaining *identity* on the one hand, and Markedness constraints reducing *contrasts* on the other, might be expected to result in a slow erosion of contrast between lexical items over the diachronic evolution of a language.

Dispersion Theory (Flemming 1995, Padgett 2001, 2003) introduces a class of Markedness constraints within OT that are directly sensitive to systemic contrast. A constraint in this class is violated when a specified contrast between any lexical items within a system falls below a specified

---

inevitably be lost through language change. Otherwise, we would expect to bound; if this dispersion constraint is ranked above Faithfulness, change will be introduced somewhere within the system to re-establish contrast sufficient to satisfy the constraint. Dispersion constraints have been used, for example, to model push-chain vowel shifts (Flemming 1995) and palatalization contrasts (Padgett 2001, 2003).

If we pursue the hypothesis that there are grammatical processes that have some causal connection to contrast, we are by no means obligated to propose that the grammar system itself be directly sensitive to contrast *per se*, as in a number of accounts of grammatical contrast maintenance (e.g., Martinet 1955, see Blevins 2003: 304f for a general discussion of long-range teleology in phonology). Rather, a relation between the synchronic grammar and contrast-maintenance can also emerge if the relative contrastiveness of output forms exert some indirect influence on the *diachronic* evolution of lexical and/or grammatical patterns.

Just such a mechanism for a system's history to indirectly influence its present form is provided by the well-established Darwinian paradigm of selection and differential inheritance. If we begin with the assumption that lexical patterns develop though negotiation (e.g., Saussure 1916 (1959),

---

find some languages that consist of just a small number of homonyms.

Labov 1994, Kirby 1999, Worden 2002) within a speech community over many cycles of communication/transmission, any mechanism in the communicative process that gives a selective advantage to easily understood (i.e., easily categorized) forms relative to less easily understood forms will promote distinctiveness between different forms, asymptotically pushing the system toward the point at which failure to categorize a form reaches some stable minimum.

The idea that linguistic contrasts may be indirectly maintained through selection and differential inheritance operating over utterances is not by any means new. In the context of phonology, this idea has been cogently articulated in Lindblom, MacNeilage and Studdert-Kennedy (1984), Lindblom (1986, 1992, 1998), Labov (1994) and Guy (1996) but independent proposals can also be found in Keller (1984 (1994)) for historical linguistics, and Hurford (1987) for syntax. In the last decade, a great many simulations of linguistic processes have been constructed that incorporate differential inheritance in some form as a guiding mechanism for pattern formation (e.g., Briscoe 2002, Cangelosi and Parisi 2002). Finally, these ideas can be seen as part of a broader recognition that the basic mechanisms of Darwinian evolution are not limited to influencing change within biological species, but rather apply to any dynamical system exhibiting variation among elements, selection of variants over some criterion, and subsequent reproduction (Dawkins 1983, Dennett 1995). Under the view promulgated here and by many others (e.g., Lindblom 1986, Ohala 1989, Labov 1994, Hurford 1999, Cziko 1995, Bybee 2001) language is such a system: a broad variety of factors result in stochastic variation between instances of 'the same' utterance, and some of those variants are more likely to be understood, learned, and imitated than others, thereby producing the differential inheritance that is the cornerstone of Darwinian evolution.

This chapter will explore the idea that differential selection and inheritance of forms on the basis of categorization success can serve as a counterweight against the tendency to uniformity in an analogically evolving lexicon, resulting in the maintenance of a functional degree of categorial contrast. To model selection of forms on the basis of categorization, the program underlying the simulations presented in this chapter contains a pair of speaker/hearers that take turn communicating their contents to each other, rather than just one lexicon that reproduces its contents for itself. In a given round, one of the pair utters all of its lexical items for the other, who attempts to recognize the utterance through comparison to the contents of its own lexicon (Luce and Pisoni 1998, de Boer 2001). Upon hearing an utterance, each lexical category in the hearer's lexicon becomes 'activated' in proportion to its similarity to the utterance, allowing lexical categories to compete for recognition. The probability that an utterance will be successfully recognized is greater to the extent that there is a single lexical category that is highly activated; if no categories are activated at all, the utterance may fail to be recognized. On the other hand, if a number of competing categories are activated to a similar degree, the utterance may be randomly assigned to one of the activated categories. When an utterance is successfully matched to a lexical category, it is stored as an example of that category, replacing a previously stored exemplar at random[16]. A more detailed description of the simulation architecture follows below in section 2.3.

Pattern evolution within the general model presented in the previous chapter proceeds because remembered tokens of a sound-category influence future pronunciations of that, and other similar sound-categories, thereby creating feedback loops between cycles of speaking and hearing which allow sound-categories to shift their centers and boundaries over time. In this way, a portion of the effect exerted by percepts on lexical evolution depends on their patterns of storage as tokens of particular categories.

---

[16] Note that successful matching and storage does not require that the match be to the same lexical category intended and encoded by the speaker.

Crucially, consistent categorization depends not only on *similarity* of a percept to a category, but also on *dissimilarity* to other categories (see Pierrehumbert 2001a for discussion of this point in the context of exemplar models). More specifically, a percept can present difficulty in categorization either when it lies far outside the boundaries of previously established categories (as in a non-native sound), or when it lies near the boundary between two categories (as in American English, for example, where nasalized [E®] lies near the boundary between [E] and [I]). In this model then, if low contrast reduces the overall probability that a percept will be assigned to a given category, low contrast percepts will make up a smaller proportion of the regularly stored tokens in that category than those that are more contrastive, i.e. more nearly uniquely categorizable. In the end then, percepts that are either outliers (lying far outside the normal featural ranges), or low-contrast (lying too near the boundary between two categories) will play a more diffuse role in guiding the future evolution of the system, through their lower rate of consistent storage as tokens of any given category (Guy 1996).

To illustrate the feedback connection between production and perception and its role in producing categorical contrast, consider recent work by Pierrehumbert (2001a, 2002, 2004) studying the evolution of phoneme categories under the influence of the feedback loops inherent in exemplar models when production and perception interact. In Pierrehumbert's perception model, the phonetic characteristics of an incoming stimulus serve to locate it in the parameter space, at which point nearby exemplars are activated according to similarity and resting activation level. Activation is passed up to the category labels, and the stimulus is then assigned to the category label with the most aggregate activation.

Production proceeds in the opposite direction, beginning with activation of a category label (e.g., Levelt 1989). The selection of the phonetic target is modeled as random selection and subsequent averaging of a set of exemplars associated with that label, where the likelihood that a particular exemplar will be selected is proportional to its resting activation level. After selection, the average is passed on to phonetic implementation, which introduces a small amount of noise in the process of output production.

Initially, a category is seeded with a single exemplar, and the simulation begins with activation of that category followed by production of that exemplar. The simulation then processes the token just produced, and if it is perceptually distinguishable from the exemplar in memory, a new exemplar is stored under the same label. In the course of simulating the development of a single category, the single seeded exemplar gives rise to a Gaussian distribution of variants under the influence of noise in phonetic implementation. The variance of the distribution increases rapidly at first, but because the production outputs represent a consensus over a random set of exemplars from that category, the output is the product of blending inheritance. Hence, the distribution shows reversion to the mean, resulting in the eventual stabilization, or entrenchment, of the category[17]. The related 'Schema Abstraction' categorization model (Hintzman 1986, Goldinger 1996) also incorporates a form of blending inheritance to counter the tendency of categories to spread under noise. Where in Pierrehumbert's model a random set of exemplars from a category is averaged to produce a production target, in the Schema Abstraction model it is the stages of perception and storage that proceed through averaging of multiple forms: in the latter model, a percept activates a set of exemplars in memory, producing an aggregate 'echo' of activation. An average over this activation pattern is stored in memory as a new 'exemplar' of the percept; the resulting blending of forms in storage results in entrenchment of the category.

---

[17] Pierrehumbert goes on to show that introduction of directed bias in the noise of phonetic implementation results in gradual shift of the phoneme category in the direction of the bias, modeling phonetic drift under the influence of articulatory markedness.

Redford et al.'s (2001) model of the emergence of canonical syllable-type inventories also proceeds through a form of blending inheritance, in which in each cycle vocabulary items 'cross-over', exchanging sequences with one another to form mixed forms. The resulting forms in each cycle are then subjected to selection on the basis of a number of functional criteria. This continual blending and selection results in a steady pressure for vocabulary items to become more alike. However, one of the functional criteria included in the selection is that vocabulary items must be recognizably distinct from one another, preventing vocabularies from devolving to full homophony.

As argued above in section 1.6.2.1-3, any form of blending inheritance in transmission, at any point in the production/perception cycle, will result in reversion to the mean over those forms. In the simulations presented in this dissertation, blending of forms, both within and across categories, is modeled as occurring in production through cross-category influence in target assembly. Note however that this choice is made for computational simplicity only – no claim is made that cross-category influence, or other mechanisms producing blending inheritance cannot operate in perception as well, as for example, in the perceptual magnet effect (Kuhl 1991).

Since Baudoin de Courtenay (1895 [1972]), linguists have proposed that factors in the linguistic environment give rise to directional, gradient biases in the form of utterances, which ultimately serve as the raw material for grammaticalization (see e.g., Ohala 1981, 1989, Ladefoged 1984, Bybee 1985, Kiparsky 1995, Bybee 2001, Blevins 2003, for more recent arguments in favor of this position). Pierrehumbert's integrated production-perception exemplar model of sound categorization provides a clear mechanism for the link between perceptible phonetic detail, and the possibility of grammaticalization as the basis for a category. Because the criteria for category membership emerge from previously categorized exemplars, any gradient feature that happens to be disproportionately present in many exemplars of a category relative to exemplars of nearby categories will increase the probability that some new stimulus exhibiting that feature will be likewise categorized. This in turn leads to a greater probability of producing an example of that category with that feature, and so on. The resulting feedback loop accelerates as the proportion of exemplars in a category showing a feature approaches 100%, providing a mechanism for phonetic features frequently associated with some category to occasionally rapidly shift from gradient, optional reproduction in implementation, to categorical, mandatory inclusion.

These results fit well with the finding that correlation alone is sufficient to initiate categorization (Maye and Gerken 2000), suggesting that any phonetic details, if perceptually salient and sufficiently correlated with some other utterance feature, can become part of the inclusion criteria for some category. Note that the probability that a feature will become categorically associated with a category is dependent on its association with that category to the *exclusion* of neighboring categories. This model therefore allows for the occasional rare association of phonetic features in a category initiated by random distributional fluctuations, but predicts that most categorical associations will derive from biases present in the environment that promote those particular categorical associations to the point that they become more reliable (e.g., Albright 2002).

In this chapter, we will see that the interaction of cross-category pressure toward uniformity on the one hand, and selection pressure for contrast on the other, can account for several general observations about lexical contrast systems. Before introducing these observations, however, it will be helpful to go over some terminology. In referring to the properties of phonological contrast systems, I use the term *contrast space*, which refers to a multidimensional space described by all possible combinations of potentially available contrasts. Note that the same effective space may be described from different points of view using different combinations of dimensionality and granularity; for example, from an acoustic point of view, vowel contrast space is often operationally described by the two dimensions of F1 and an F2 / F3 combination (e.g., Lindblom 1986). Alternatively, the same vowels could be described articulatorily within a contrast space comprising the three dimensions of height, backness, and rounding. In the acoustically defined, two dimensional contrast space, more points along each dimension are required to describe a given set of vowels than in the three dimensional articulatorily defined space.

With this terminology in hand, I turn to the three general observations about phonological contrast systems that we'll investigate in the simulations in this chapter:

1. Phonological systems do not appear to ever evolve through states in which most or all categorial contrasts are lost.
2. Phonological contrast systems tend to make use of symmetrical inventories, rather than widely scattered contrasts (e.g., Lindblom et al. 1984, Lindblom 1992).
3. When a marked element appears in a system of contrast, its frequency is usually lower than that of less marked system elements (Greenberg 1966).

In sections 2.4-6 below, I discuss each of these observations in more detail, and show that when differential selection and inheritance based on categorization are introduced into simulations of analogically evolving lexicons, structures emerge that exhibit these properties. Before we get there though, section 2.3 introduces the simulation architecture employed here.

2.3 Architecture of the simulation

The program showcased in the previous chapter includes no mechanism for evaluating and responding to contrast between lexical items, which had the practical consequence that only very limited kinds of change could be permitted – otherwise, all lexical items would rapidly evolve to be the same by virtue of cross-category pressure in target assembly. The simulations presented in this chapter are carried out using a more complex program that incorporates communication between two independent speaker/hearers, where the consistency with which a hearer is able to categorize a given utterance provides an additional selective parameter on evolution of word forms.

2.3.1 Brief overview of the simulation architecture

Like the simulation architecture presented in the first chapter, this simulation models the development of patterns in a lexicon that evolves through the interaction of cross-category pressure within the lexicon, and performance biases external to the lexicon. In the simulation, two speakers each possess a small lexicon containing a set number of categories (see Figure 22 in section 2.3.8 below for a schematic diagram of information flow in the simulation). These categories are structured as exemplar clouds (Goldinger 1996, Johnson 1997, Pierrehumbert 2001a), where each lexical category has a fixed number of exemplars stored from past experience. Each exemplar is coded as a string of segments specified through feature-bundles; 4 segments are allowed in the simulation, named for convenience [k, x, i, a]. The four segments are characterized by two features, for convenience named 'height' and 'voice'. There are four height values, with [k] being the highest and [a] the lowest; [k, x] have a voicing value of 0, while [i, a] have a positive voicing value. At the beginning of a simulation, the lexicons of each speaker are seeded with a full complement of exemplars.

As before, cross-category pressure is modeled at the level of target assembly, where lexical entries that share substrings with a given entry are able to contribute to its production, resulting in an output that is occasionally biased toward other similar forms. Likewise, beyond this effect of cross-categorization in the lexicon on target assembly, the actual production of an assembled target may be further biased through performance factors external to the lexicon. No provision is made for the addition or deletion of segments, with the result that exemplar string length remains constant throughout a given simulation.

A major difference between the two architectures appears, however, post-production. In the simulation architecture used in Chapter 1, a produced exemplar is simply re-stored in its lexical entry of origin, with no concern for how similar that exemplar is to other exemplars stored under that lexical entry, or to any other exemplars stored elsewhere. In the simulation presented here, on the other hand, categorization is modeled as a step in the cycle of production and re-storage of lexical items. To introduce a role for categorization, this program incorporates two independent speaker/hearers; in a given round one takes the speaker role and utters the contents of its lexicon for the other, who attempts to understand, that is to say, categorize, the utterances. Categorization proceeds by comparison to existing exemplars in the hearer's lexicon. If an incoming utterance matches the exemplars of one category well, but few others, it will be assigned to that category with high probability. If it matches nothing well, or is a good match to exemplars of more than one category, it will fail be consistently assigned to a particular category. Matched utterances are stored in the hearer's lexicon as an exemplar of that category. Storage of a perceived string in a lexical entry results in replacement of a randomly chosen, previously stored exemplar from that entry, resulting in a slow turnover of exemplars in the lexicon.

In this model, stored exemplars serve as models for future production events, based on evidence that production can be influenced by temporally distant, phonetically detailed memories of percepts (Goldinger 2000). To show this, Goldinger had a group of subjects produce a baseline recording of a set of words. The next day they *heard* the same stimuli words spoken some number of times in a particular voice and were given the task of identifying that word in a grid. Five days later, they returned and made test recordings of the same list of words again. For each word recorded by each subject, an AXB test stimulus was made from (i) the subject baseline recording of the word, (ii) the word as heard by the subject the second day, and (iii) the final subject test recording of that word (where the order of the baseline and test recording were random). These recordings were played for a separate group of listeners, who were given the task of rating which of the two subject recordings of the word was more like the middle word in the other voice. The success of listeners in identifying the second test recording as more similar to the reference was well above chance, indicating that for the subjects, phonetic details of a pronunciation heard five days earlier had a significant influence on their own production of that word. The correlation was higher when the word was repeated more times in the identification task on the second day.

This recency effect of perception on production is incorporated into the model by allowing previously stored exemplars to be steadily replaced by incoming perceived exemplars, giving more recently stored exemplars a greater chance of serving as a current model for production. Because stored exemplars serve as models for future production events, forms that are consistently stored as members of a given category have a greater influence on that category's evolution than forms that are less often stored in that category. This selective differential obtains because an easily, uniquely matched form contributes more consistently to a single category's future outputs, while the influence of a poorly or promiscuously matching form is more diffusely distributed over the lexicon. Because utterance forms are traded back and forth between the two speakers, a feedback loop is closed in which utterances with more influence on future production events are more likely to persist in the system than those with less influence.

It will be demonstrated in section 2.4 below that in the context of this feedback loop, differential consistency in categorization results in selection on the basis of contrast, with the result that categories tend to evolve to contain forms that are sufficiently distinct from one another that their outputs are nearly always correctly categorized by a hearer. Importantly, note that there is nothing in the simulation that assesses contrast directly. Contrast only plays a role in the consistency of category assignment, which in the context of the feedback loop between production and categorization indirectly results in the evolution of contrastive categories. The simulation architecture is described in greater detail beginning in section 2.3.2 below.

2.3.2 The structure of the lexicon

In the previous simulation, lexical entries consisted of an abstract category label, and an associated lexical entry which was split into an invariant input level , and an evolving exemplar-set level. Generalizations over the lexicon that influenced target assembly were then based on the mappings from each permanent input to its associated exemplars. In that system, the anchor of the permanent input level was required in order to prevent lexical entries from collapsing into uniformity under the force of cross-category pressure.

In the current simulation in contrast, a lexical entry consists solely of an abstract category label and associated stored exemplars, with no underlying permanent input level. As will be demonstrated below in section 2.4, the selective advantage of contrast in output forms in this simulation architecture counterbalances cross-category pressure toward uniformity, resulting in preservation of adequate contrast between distinct lexical items.

Each lexical entry is also associated with a token frequency, which determines the number of times that lexical entry will be produced in a given round. In the simulations presented here, each lexical entry contains a set number of exemplars that is three times the token frequency; when a percept is categorized as an exemplar of a given lexical entry, it replaces a random exemplar in that entry, such that the total number of exemplars in each entry remains the same over the course of a simulation. Because the number of stored exemplars in the lexicon is fixed at three times the number of tokens of that category produced in every round, total replacement of the contents of a lexical category requires, at minimum, three rounds of listening within the simulation, providing a certain degree of stability and inertia in lexical categories.

2.3.3 The structure of exemplars

As in the first simulation, lexical entries are structured as strings of segments that are characterized by sets of feature values. Again, these values are solely meant to serve as cues that can be used for categorization within the simulation, and are not to be taken as a crucial endorsement of one or another theoretical system of phonological contrast. However, to provide a structuring metaphor for labeling these features and conceptualizing their interactions, the architecture of the simulation borrows Gestural Phonology's (Browman and Goldstein 1989) conception of features as articulators, and feature values as articulator target positions.

Phonological structure of simulated lexical entries begins with assuming the existence of segments, i.e. temporally ordered sub-groupings of features within entries. Feature labels are conceived as identifying a particular articulator, while feature values represent a particular target position for that articulator within the segment. For example, if the 'hi' feature is taken to refer to the tongue body, then a feature value for 'hi' of 0.3 out of a possible range of 0 – 0.3 corresponds to a target of full closure for that gesture in that segment. A value of 0 in contrast refers to a position at the opposite extreme of the potential range. As in the previous simulations, performance biases can intervene in certain contexts to make it more or less likely that a given target will in fact be reached. Edges of lexical entries are marked by an extra segment slot containing the 'feature' [LeftEdge] or [RightEdge], respectively. No relative timing of gestures within or between segments is assessed.

In all the simulations that will be described below, each lexicon consists of up to 15 lexical entries, each lexical entry exemplar consists of 3-4 segments, and each segment consists of two articulator features, which can each take a range of values on a scale. For computational simplicity, the scales are coarsely granular; specifically, each articulator is limited to a range of 4 or fewer values. This restriction produces de facto feature categories that can be easily compared to one another within the simulation; in effect, the simulation does not extend the available levels of categorization below the level of feature-value strings to feature-values themselves (see Pierrehumbert 2001a for an example of categorization at the level of feature-value). When categorization is extended to this level in some future version of the program, more continuously varying feature-values can be investigated; the results presented in this dissertation suggest that more continuously-varying features should also come to cluster together in identifiable categories through the same cross-category mechanisms that coerce feature-value sequences to cluster.

The sole pathway for change in the simulations presented here lies in variation in a feature's value. Within target assembly, a feature may change through substitution of one value for another through influence of other similar sequences in the lexicon, and in performance, a feature's value may be incrementally increased or decreased through random variation or performance bias. There is no mechanism for varying timing relationships between segments, and therefore there is no way for a segment to grow shorter and eventually delete through a failure to be consistently perceived. Likewise, the simulation does not yet allow any mechanism for interpolation or substitution of a segment. Therefore, the length of the exemplars within a lexical category remain fixed over the course of the simulation.

Only two features are included per segment in the simulations shown below,
[hi] with the four allowed values.0, .1, .2, .3, and [voi] with the two allowed values .0 and .1. Furthermore, the values of [voi] are pegged to the values of [hi], such that the low closure values of [hi], [hi: .0] and [hi: .1] obligatorily have the feature values [voi: .1], while the high closure values, [hi: .2] and [hi: .3] obligatorily have the feature values [voi: .0]. For ease of reading the results of simulations, the following shorthand is used:

[hi: .0], [voi: .1]:	a
[hi: .1], [voi: .1]:	i
[hi: .2], [voi: .0]:	x
[hi: .3], [voi: .0]:	k

By pegging the value of [voi] to ranges of values of [hi], we allow the program to develop patterns based on a class of segments as well as based on individual segments alone. An alternative way to think of this is that the [voi] feature serves as a way to divide the inventory into larger C and V categories. This will be discussed in more detail in section 2.6 below.

The four-segment contrast-space available for exploration by an evolving lexical category in these simulations is quite dramatically smaller than that available to a lexical category in a natural language. However, as will be seen in sections 2.4 through 2.6 below, the range of contrast space is nonetheless sufficient to begin exploring the issues set out in this dissertation.

2.3.4 Target Assembly

To briefly recap discussion from section 1.6.2.1-3 in the first chapter, this dissertation explores the idea that lexical characteristics that are highly correlated with one another form basins of attraction in the course of production, resulting in a general warping of output forms further toward the centers of those basins. A possible mechanism for this warping emerges if target assembly proceeds through spreading activation of many different levels of categorization (Pierrehumbert 2001a, 2002, 2004, McLennan, Luce and Charles-Luce, 2003). Since every incipient output form will contain features and feature sequences that are cross-categorized with many other forms in the lexicon, target assembly will tend to reproduce correlations that are present in the lexicon as a whole. 'Cross-category pressure' is the cover term used here to refer to any mechanism allowing blending inheritance, and thus entrenchment of categorical behavior in output form over the lexicon as a whole. Additional non-exclusive processing pathways producing blending inheritance are plausible: for example, in matching percepts to categories, conflicting top-down influences on perception may warp perception toward the centers of adjacent categories (Kuhl 1991). Alternatively, in articulation, pattern reinforcement through the entrenchment of motor programs in proportion to degree of practice may plausibly contribute to patterning at the scale of the segment and segment transition (Saltzman and Munhall 1989, Shadmehr and Bashers-Krug 1997, Bybee 2001, Ussishkin and Wedel 2003, Wedel 2004).

As in the previous simulation, the effect of spreading activation of overlapping/nested categories on target assembly is modeled through the identification of similarities between a provisional target and a sampling of other sequences present within the lexicon, and through provision of a mechanism for similar sequences to influence the form of the final target. Target assembly for a given lexical entry proceeds therefore not only through reference to previously stored exemplars in that entry alone, but also to previously stored exemplars in other lexical entries, modulo feature and feature-sequence similarity.

As before, the program detects patterns in the lexicon by finding all feature matches between a reference string within the provisional target and a comparison string, and then writing a generalization for each possible combination of those feature matches. As discussed previously, writing generalizations on the basis of all contentful sets of matches (i.e., the power set of matches minus the empty set) allows those sets of feature values that are reliably or frequently associated with one another to be straightforwardly identified within the simulation. Where the previous program placed no limitation on the lengths of segment strings to be compared, the program used in this chapter compares strings of at most two segments in length, because the decomposition of segments into several features produces large power sets in a fewer number of segments. As in the first chapter's program, it bears repeating that the mechanism employed by the program to detect similarity between exemplars is not intended to make any specific predictions about the ways similarity might be computed in human language production or processing.

In the previous chapter, it was found that factoring the effect of the generalization set on outputs by either frequency, reliability, or by Skousen's ALM algorithm all produced similar results. In this simulation then, I have chosen to focus on reliability of generalization, with the caveat that generalizations that hold over very few forms in the lexicon, even if highly reliable, form a shallower basin of a attraction than generalizations holding over many forms (see Albright and Hayes 2002 and Albright 2002 for a detailed discussion of this approach within the similar Minimal Generalization framework).

Assembly begins with selection of a lexical category for production, and selection of a random exemplar from within that category. This reference exemplar forms the basis for target assembly. For each non-edge segment, starting from the left, the two two-segment spans comprising this reference segment and its neighbor to the left, and then its neighbor to the right are compared to all two segment spans within a randomly chosen subset of exemplars from every lexical category in the lexicon. For all contentful combinations of matches between the reference span and each comparison span (the power set of matches minus the empty set), a generalization is written with the match-set as antecedent, and the feature values of the corresponding segment in the comparison span as consequent. These generalizations will be used to detect correlations between features and feature values within the lexicon, but before describing this further, let's look at an example of the process so far. An example of the generalizations deriving from the comparison of a reference segment at the left edge of a reference string to another string is shown below in Figure 22. In the moment in the simulation shown here, the first slot of an exemplar chosen as the basis for target assembly is being compared to the first slot of some other exemplar stored in the lexicon; comparison includes comparing the slots to the left and right of the current slot (see Fig. 22A). In this case, the algorithm notes that there is a match in the slots to the left, [LEdge], a single match in the central slots, [voi: 0.0], and a match in the slots to the right, [hi: 0.2]. All combinations of matched features are computed that contain up to two contiguous slots. These combinations are written in Fig. 22B. Each of these combinations is treated as a potential generalization, and the entire feature specification of the central slot in the comparison exemplar is added as a consequent for each of the antecedents (Fig. 22B). After this process is complete, the comparison string (exemplar) is shifted one slot further to the right with regard to the reference string, and the process is repeated. After this has been done for every exemplar in the lexicon, we will have a set of antecedents that have share some portion of the context for our current reference slot, and associated with each antecedent is a set of consequents that tell us what, in the lexicon, is associated with that particular context, as described below.

Figure 22. Comparison and Generalization

A. Current reference segment

Current Reference string: [LEdge], [hi: .2, voi: 0], [hi: .2, voi: .1]

Matches:

Current Comparison string: [LEdge], [hi: .3, voi: 0], [hi: .2, voi: 0]

B. Generalizations added for the first segment in the current reference string:

	Antecedents			Consequents
1.	[LEdge]	—		[hi: .3, voi: 0]
2.		[voi: 0]		[hi: .3, voi: 0]
3.		—	[hi: .2]	[hi: .3, voi: 0]
4.	[LEdge]	[voi: 0]		[hi: .3, voi: 0]
5.		[voi: 0]	[hi: .2]	[hi: .3, voi: 0]

Whenever a new antecedent is the same as that of a previously written generalization, rather than writing an entirely new generalization, the consequent (the feature-value set of the central reference span segment) is simply appended as an additional consequent to that antecedent. In the example above for instance, a previous generalization with the same antecedent as '1.' above,

"[LEdge]   —"

might already exist in the list of generalizations, with, say, the following associated list of consequents:

[hi: .1, voi: .0]

[hi: .2, voi: .1]

[hi: .1, voi: .0]

[hi: .0, voi: .0]

[hi: .2, voi: .1]

In this case, the feature-values of the current comparison span, "[hi: .3, voi: .0]", will be added to this list of consequents. After all generalizations have been created, each generalization is assessed for the reliable appearance of a given feature-value in the consequents. In the example above, the consequent list for the antecedent, “[LEdge] — “, does not show any particular pattern, indicating that a feature value of [LEdge] does not make any particular prediction about feature-values in the subsequent segment in this particular lexicon.

Imagine however that another lexicon has been evolving with a bias in performance against high-sonority segments in word-initial position, such that most word initial segments in the lexicon have evolved to show high closure. Words that contribute to the generalization with the antecedent matching (1.) above,

“ [LEdge] — ”

will therefore show a skewed set of consequents, say:

[hi: .3, voi: .0]

[hi: .3, voi: .0]

[hi: .3, voi: .0]

[hi: .2, voi: .0]

[hi: .3, voi: .0]

where the value of ‘voi’ is very consistent, as is the value for ‘hi’.

In the simulation, each feature position in a generalization consequent list is checked for consistency, or *reliability*, by calculating the standard deviation of the feature values. When most feature values at a position are the same or very similar, the standard deviation is low. In the simulations presented in this chapter, the reliability required for a pattern in the lexicon to have any influence on target assembly is set at a standard deviation of 0.03. In the case where the consequent list contains at least three examples, (otherwise the generalization would hold over very few forms), the program may randomly chose a value from the consequent list at that feature position to appear in the final assembled target at that position. For example, then, in a lexicon where the bias against low-sonority word onsets has resulted in a widespread loss of closure contrast in this context, even if the provisional target happens to contain a low-closure initial segment, the initial segment is quite likely to emerge with high closure in the final assembled target. In this case, the high reliability of the feature-value [hi: .3] following [LEdge] in the lexicon may lead the target assembly process to replace the reference feature-value [hi: .1] with [hi: .3].

2.3.5 Performance

After target assembly, the string of feature-values is passed to Performance. As in the simulations described in the first chapter, no distinction is made within the program between biases that are thought to arise in articulation versus perception or any other source outside the lexicon for that matter. As before, performance both introduces small random variations in each round, and also administers specific biases, which can be set individually in each simulation, that identify certain combinations of articulators, values, or positions within an assembled target (e.g., low closure at the left edge). When a specified context is identified within the assembled target, the bias may intervene and increase or decrease the value of a specified feature within that context.

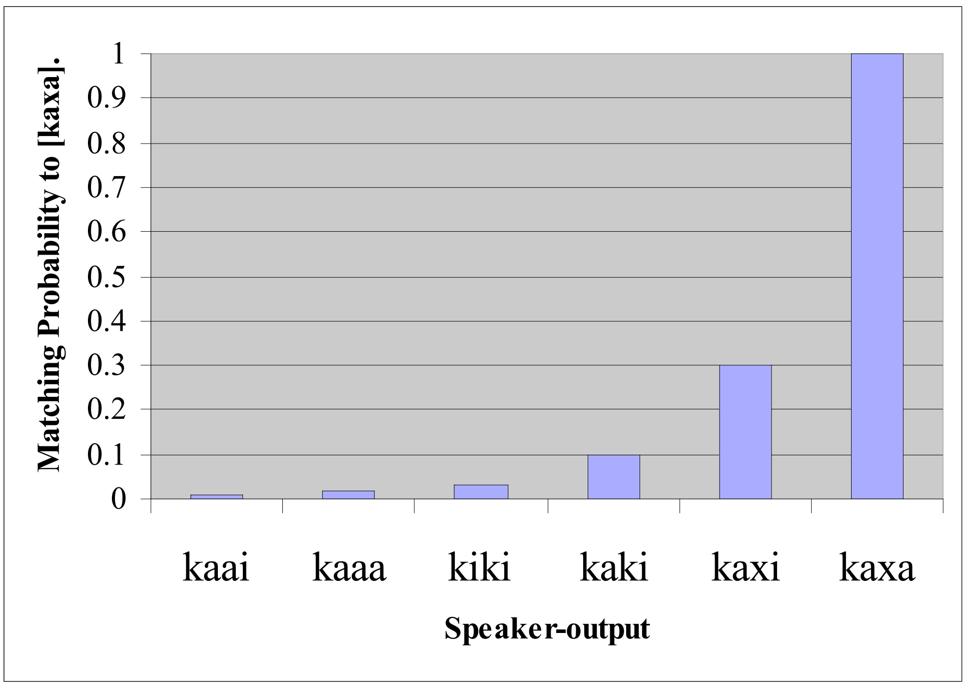
Because this study is not concerned with pattern formation associated with the processes of translating between articulatory targets in production and acoustic signals in processing, the program in its present form does not model this back and forth transformation (e.g., Goldinger 2000), but rather makes the computationally simplifying move of leaving all information in the form of articulatory targets as described above (e.g., Johnson 1997).

2.3.6 Lexical Access and Categorization.

The hearer receives an output from the speaker, modified through the filters of Performance as described above. Matching an incoming form to lexical categories is modeled in the following way. Starting from the left, the listener attempts to match each segment of the incoming speaker-output with all exemplars within its lexicon. Matching is stochastic, such that near-matches will sometimes be registered as matches. A running match-tally is computed for each listener-exemplar as matching proceeds; if the final proportion of actual matches to the total possible matches falls below a set value, the exemplar under consideration is removed from the pool of potential matches (for conceptually similar approaches, see Nosofsky 1989, Pierrehumbert 2001a).

The probability that a near-match will be registered as a match is determined both according to degree and number of mismatches. For example, a speaker-output that is different from a potential match in the hearer's lexicon at a distance of 0.1 in a given feature will be registered as a successful match ~30% of the time, while a speaker-output that differs by 0.2 will be registered as a successful match ~2% of the time. The probabilities that an set of example speaker-outputs will be registered as matches to the hearer-exemplar [kaxa] are shown below in Figure 23.

Figure 23. Probability of matching to [kaxa]



The speaker-output [kaai] differs from [kaxa] by .2 in the third segment and by .1 in the fourth; [kaaa] differs by .2 in the third segment; [kiki] differs by .1 in the final three segments; and so on. No significance is attributed to the particular method of assessing likelihood of matching; altering the shape of the matching function alters the minimum stable distance between categories, but does not qualitatively alter the results presented below (not shown).

When an exemplar is successfully matched to the incoming speaker-output, the label of its lexical category is added to a growing set of competitors for identification. Because every exemplar is examined from each lexical category, the label of a lexical category whose exemplars match a speaker-output closely will appear more often in the set of competitors than that of a category whose exemplars match the output less well. For example, if the incoming speaker-output matches most of the exemplars in lexical category A perfectly, but also matches some of those in lexical category B more poorly, the final list of competitors might look like, "A, A, B, A, A, A, B, A", for example.

After every exemplar in the hearer's lexicon has had its chance to be matched, the program allows two distinct procedures to be applied to determine how an incoming speaker-output is identified and stored, which we might call 'conservative' and 'liberal', respectively. In the conservative version of lexical access, an incoming speaker-output must satisfy two constraints in order to be recognized and stored as a new exemplar of a given lexical category: it must both be matched to some minimum number of exemplars in the lexicon, and further must have matched that category significantly more often than others. In the program, this is accomplished by requiring that the list of competitors be above a minimum length, and then by randomly choosing (with replacement), say, four members of the list and then requiring that at least three of the four be the same. Given the list "A, A, B, A, A, A, B, A", it is clear that many, but not all sets of four randomly chosen set members will contain at least three instances of A, and occasionally, the set will contain three instances of B. When the majority criterion is met, the incoming speaker-output is stored in the hearer lexicon as a new exemplar of that lexical category, in place of a randomly chosen previously stored exemplar. If either of these criteria are not met, the incoming speaker-output is not stored at all. In the conservative procedure then, an incoming speaker-output must not only be similar to a minimum number of exemplars in the hearer lexicon, but matching is more likely to the extent that those exemplars are primarily from one category.

In the liberal lexical access procedure, if an incoming speaker-output has succeeded in activating anything at all in the hearer's lexicon beyond a certain threshold, it will be stored. The list of competitors is not checked for minimum length, and an identification is made simply by choosing a category label at random from the set of matches. The liberal and conservative procedures are similar, however, in that an incoming speaker-output is likely to be identified with a particular category in proportion to how well it matched exemplars from that category, and how poorly it matched to others.

These different procedures are intended to coarsely approximate two relatively extreme possibilities for processes of competition and categorization within the hearer's lexicon as an incoming speaker-output is processed, each of which provides a selection against less contrastive forms. Under the liberal lexical access procedure, the matching process ensures that any incoming speaker-output that remotely matches anything in the hearer's lexicon will be stored somewhere. However, speaker exemplars from a poorly contrasting lexical category will tend to split their contributions between several categories in the hearer's lexicon over time, while speaker exemplars from a more contrastive lexical category will tend concentrate their contributions on one category in the hearer. Under the conservative procedure, outputs that either do not match any category well, or that are good matches to multiple categories, fail to be stored at all.

2.3.7 Deixis

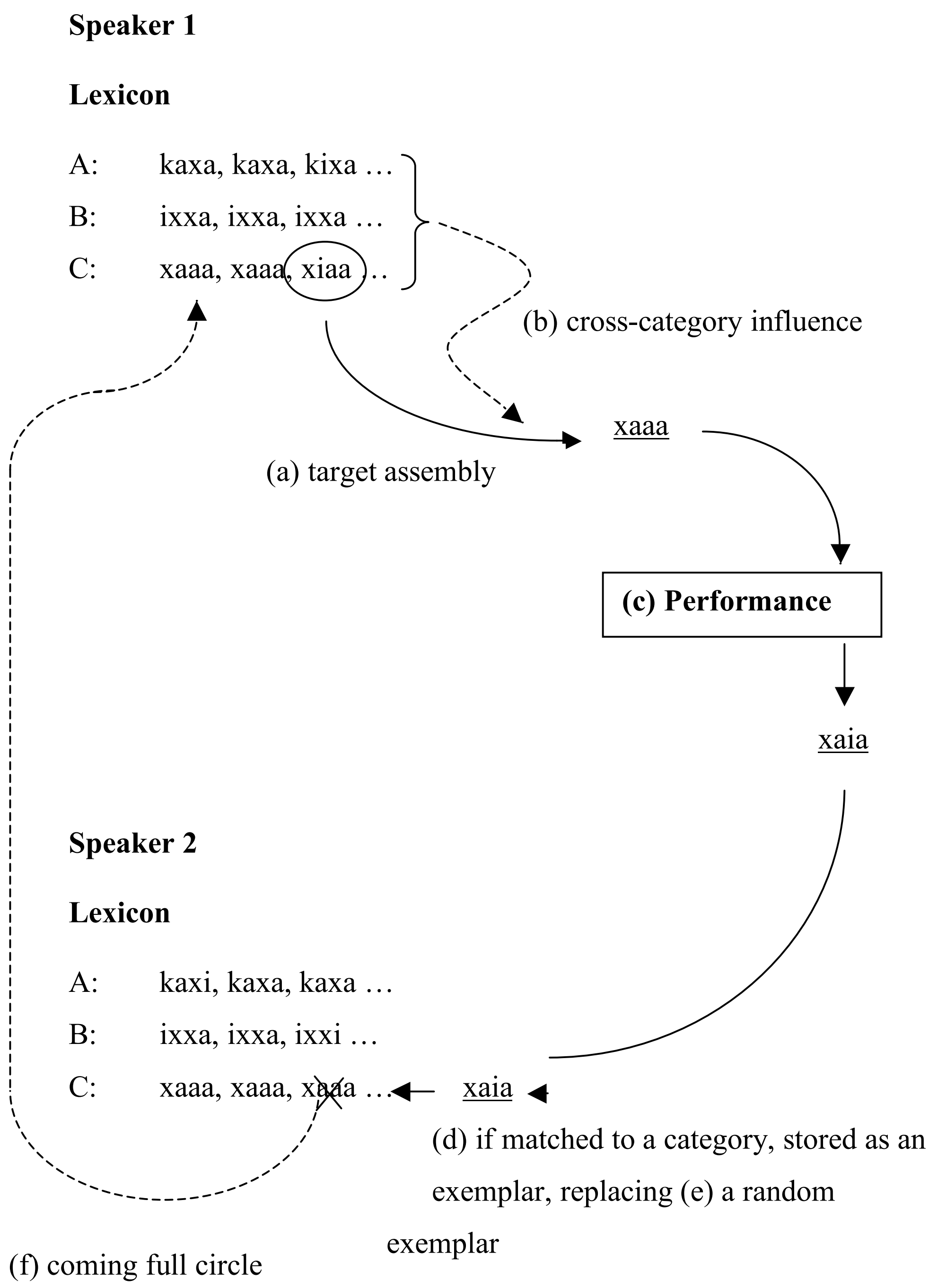
Under either the conservative or liberal lexical access procedures, the speaker-hearer pair rapidly develop a jointly held system of lexical categories as will be shown in section 2.4 below. However, in the absence of any external context to ground the connection between a feature string and a category label, the pair may come to associate distinct labels with a given feature string. The two lexicons can be brought into correspondence by allowing the speaker to bypass hearer lexical access in a few percent of utterances by simply 'showing' the hearer directly what category was intended by the exemplar (e.g., Oliphant and Batali 1997). In the absence of such 'deixis', the pair still quickly negotiate a jointly held set of words that each can categorize, but the word-label correspondences may be different. Because the results shown in this chapter do not rely in any way on the two speakers agreeing on the category label for a given set of exemplars, all storage was mediated solely through categorization, i.e., with no deixis, in the simulations shown in this chapter.

2.3.8 Cycling and summary

At the beginning of a simulation, the two speaker/hearers are both provided with identical lexicons, where each lexical entry contains three times the number of stored exemplars as will be produced from each entry in a given round. In even-numbered rounds, one agent is the speaker, producing the prescribed number of tokens of each lexical category in the manner described above, and the other agent is the hearer, receiving each output and attempting to recognize and store it. In odd numbered rounds, the roles reverse.

Figure 24 below schematically illustrates the flow of information in the program. At the start of a simulation, a fixed number of categories is set up in each lexicon, and each category is provided with a starting set of stored token exemplars. To begin a cycle, the current speaker chooses an exemplar from a lexical category, and assembles a production target based on that exemplar (a). In simulations modeling the effects of cross-category influence on pattern development, patterns inherent in the rest of the lexicon may influence target assembly at this stage, modulo similarity (b); in simulations lacking cross-category influence in target assembly, the assembled target will be identical to the originally chosen exemplar. The output of target assembly is passed through Performance, where noise or directional biases may alter its form (c). The hearer then attempts to match the speaker-output to a category in its lexicon. If the hearer successfully categorizes the speaker-output, it is stored as a new exemplar of that category (d), replacing a randomly chosen previously stored exemplar (e). In some future cycle when speaker/hearer roles have reversed, this exemplar may serve as the basis for assembly of a target from that category. If the resulting output is categorized and stored in its originating category by the hearer in that round, the information contained in the original exemplar will have come full circle (f). In the following sections we'll see how the various influences and filters on this information flow in the program can produce patterns over many cycles.

Figure 24. Summary of Simulation Information Flow



2.4 Pattern reinforcement versus selection for contrast

In the following section, I describe the output patterns that evolve in the simulation through the interaction of cross-category influence in target assembly, and selection for contrast in transmission. I begin by showing results of a simulation in which selection for contrast has been disabled and there are no biases on pathways of change in Performance, such that patterns that exist in the lexicon are the only directional source of change in utterances. Evolution under these conditions eventually results in global categoricity, that is, identity in all categories. When selection for contrast is reintroduced (section 2.4.2), local rather than global categoricity develops: categories evolve to share sequences, but contrast between categories does not collapse.

2.4.1 Pattern development driven by pattern reinforcement in the absence of selection for contrast.

The speaker's lexicons in this and the following simulations are seeded with a small lexicon consisting of four lexical categories with the labels, A, B, C and D, each containing 9 four segment exemplars that all start out with the following form:

A: [hi .3; voi .0] [hi .0; voi .0] [hi .2; voi .0] [hi .0; voi .1] = kaxa

B: [hi .2; voi .0] [hi .1; voi .0] [hi .3; voi .0] [hi .1; voi .1] = xiki

C: [hi .0; voi .1] [hi .3; voi .1] [hi .1; voi .1] [hi .3; voi .0] = akik

D: [hi .1; voi .1] [hi .2; voi .1] [hi .0; voi .1] [hi .2; voi .0] = ixax

The set of possible features-values for [hi] in all simulations shown in this chapter is [0, .1, .2, .3]. During production, every feature-value in the string under production has a 1% chance of shifting one value up or down, within the set range[0 - .3] for [hi] and [0 - .1] for [voi]. Recall that the feature value [voi]is redundant with respect to [hi] in the program: feature values of [hi] of .2 or .3 are required to have [voi] values of 0, while [hi] values of 0 or .1 are required to have [voi] values of .1. If a feature value is changed within Performance to produce a disallowed combination, the other feature is changed at random to restore a legal combination. In this and the following sections, changes introduced in Performance are random, as opposed to being biased by context. The simulation summarized in Figure 25 below used the liberal categorization procedure (described in section 2.3.7 above). Below, I show the contents of each lexical category for Speaker 1 at 100 cycle intervals up to 400 cycles.

Figure 25. Pattern reinforcement in the absence of selection for contrast.

	Cycle 0	Cycle 100	Cycle 200	Cycle 300	Cycle 400	Consensus at 400
A	kaxa	kiki	kaik	xiix	xxxx	
	kaxa	kiki	kxii	xxaa	xxxx	
	kaxa	kiki	kaii	xxak	xxxx	
	kaxa	kika	kxii	ixxa	xxxx	xxxx
	kaxa	kiki	kaik	xxaa	xxxx	
	kaxa	kaka	kiii	xikx	xxxx	
	kaxa	kiii	kaik	xiaa	xxxx	
	kaxa	kiki	kxii	xiak	ixxx	
	kaxa	kiki	kaik	xxxa	kxxx	
B	xiki	iiki	kkai	kxax	xxxx	
	xiki	kiki	kiki	kaxx	xxxx	
	xiki	iiki	kaki	kkkx	xxxk	
	xiki	kiki	kiki	axkx	xxxx	xxxx
	xiki	iiki	kxki	xixx	xxxx	
	xiki	iiki	kiai	kxax	xxxx	
	xiki	kkxi	ikik	xixx	kxxx	
	xiki	kkxi	kiii	kkkx	xxxx	
	xiki	xiki	kiki	xixx	xxxx	
C	ixax	ikik	ikka	iiax	xxxx	
	ixax	ikik	kiki	ixax	xxxx	
	ixax	ixik	aiai	iiax	xxxx	
	ixax	iaii	akki	aiax	xxxx	xxxx
	ixax	ixik	iiki	xxxx	xxxx	
	ixax	ikxi	ikkk	iiax	xxxx	
	ixax	ikik	ikkk	ikxk	xxxx	
	ixax	ikik	kiki	iiax	xxxx	
	ixax	akik	ikki	xxxi	xxxx	
D	akik	akik	ikkx	xxia	xxxx	
	akik	ikik	ikik	xxxx	xxxx	
	akik	ikik	ikak	xaxx	xxxx	
	akik	ikik	ikik	xiaa	xxxx	xxxx
	akik	ikik	akak	xaix	xxxx	
	akik	ikik	akka	xixx	xxkx	
	akik	ikik	ikia	xiia	xxxx	
	akik	ikka	xaik	xiix	xxxx	
	akik	ikik	ikax	xiia	xxxx	

There are several things to note here. First, between the 300^{th} and 400^{th} round of the simulation, the contents of each lexical category become identical, within the range of stochastic variation imposed by Performance. This evolution toward uniformity derives from the fact that whenever a featural grouping becomes by chance sufficiently correlated to cross the threshold of a standard deviation of 0.03 in the lexicon, it becomes able to influence target assembly, and from there begins to act as an ever deepening basin of attraction in assembly. Notice above that in cycle 300, there are already quite a few [xx] sequences that have appeared. Further chance alterations introduced in Performance continually produce new variations in the range of featural groupings over the many cycles of the simulation, allowing ever-larger groupings to become correlated, with the end result that the lexicon eventually becomes uniform[18]. This is, of course, just the same

---

[18] Note also that by cycle 400, every lexical item consists of the segment [x], with the result that every sequence in each individual lexical item is maximally self-similar. The lexicon evolves this way because the program is able to take into account correlations between feature-values crossing segmental boundaries up to the size of a diphone. Because the diphone sequences of a lexical item overlap, correlations within the lexicon are maximized when all segments evolve to be identical. For example in this case, given that each lexical item is [xxxx], the number of [xx] diphone sequences

phenomenon we saw in Chapter 1, Figure 3 in which dark squares eventually took over the cellular automata game, and in Chapter 1, Figure 11, in which a single syllabification took over the simulated lexicon. This is simply a further demonstration that any system in which elements can be propagated by blending inheritance will inexorably devolve to uniformity in the absence of mechanisms to re-introduce variation. Here, the only source of variation is the random change occasionally introduced in Performance (note, for example the few exemplars scattered around the lexicon at cycle 400 that do not exactly conform to the consensus 'xxxx'). The rate of introduction of random change is set sufficiently low in the simulation that at some degree of similarity across the lexicon, the rate of change toward increasing similarity becomes faster than the rate of change away from it due to random change in Performance, with the result that the global minimum state for the system is uniformity.

2.4.2 Introducing selection for contrast through competition between lexical categories in lexical access and storage.

---

per lexical item is three, whereas if each lexical item were, say, [axik], each lexical item would contain the three distinct diphone sequences, [ax], [xi] and

In the previous simulation, the hearer in each round stored the speaker-output in the intended category with 100% accuracy, regardless of the featural contents of either the speaker-output, or the exemplars in the hearer's lexicon. In the representative simulations presented in Figures 24 and 25 below, we disconnect the direct, 'telepathic' link between speaker and hearer, and make hearer categorization, and therefore storage, dependent on comparison between the speaker-output and exemplars in the hearer's lexicon. Both simulations otherwise use the same starting configuration as the simulation described above, including cross-category influence on target assembly. The simulation shown in Figure 26 employed the liberal procedure for categorization and storage (described above in section 2.3.7) while that in Figure 27 used the conservative procedure.

[ik]. This is possible in this version of the simulation because no constraints on segment sequence have been imposed.

Figure 26. Selection for contrast: liberal procedure.

	Cycle 0	Cycle 500	Cycle 1000	Cycle 1100	Cycle 1500	Consensus at 1500
A	kaxa	kika	kixa	kaxa	kaka	
	kaxa	kika	kixi	kaxa	kaki	
	kaxa	kika	kiki	kaka	kika	
	kaxa	kika	kixa	kaka	kaka	kaka
	kaxa	kika	kixa	kaxa	kaka	
	kaxa	kika	kaxa	kaxa	kaka	
	kaxa	kika	kixa	kaka	kaka	
	kaxa	kika	kixa	kaxa	kaka	
	kaxa	kika	kiki	kixa	kaxa	
B	xiki	kaxi	kixa	kiki	kiki	
	xiki	kaxi	kixa	kiki	kiki	
	xiki	kaxi	kaxa	kixi	kiki	
	xiki	kaxi	kika	kixi	kiki	kiki
	xiki	kaxi	kixa	kiki	kiki	
	xiki	kaxi	kaxa	kiki	kika	
	xiki	kaxi	kiki	kixi	kiki	
	xiki	kaxi	kixa	kiki	kiki	
	xiki	kaxi	kaxa	kixi	kiki	
C	ixax	ixak	axak	axax	akak	
	ixax	ixak	axak	axax	akak	
	ixax	ixak	axak	axak	akak	
	ixax	ixak	axak	axax	akak	akak
	ixax	ixak	axak	axak	akak	
	ixax	ixak	axai	axak	akik	
	ixax	ixak	axak	axax	akak	
	ixax	ixak	axak	axak	axak	
	ixax	ixak	axak	axak	ikak	
D	akik	akix	akix	ikix	ikik	
	akik	akax	akik	ikix	ikak	
	akik	akix	ikix	ikix	ikik	
	akik	akix	akix	ikix	ikik	ikik
	akik	akix	akix	ikix	ikik	
	akik	akix	akix	ikix	ikik	
	akik	akix	akix	ikix	akik	
	akik	akix	akix	ikix	ixak	
	akik	akik	akix	ikix	ikik	

Figure 26 above shows snapshots of each lexical category's trajectory over 1500 cycles of the simulation. There are several things worth noting here. Firstly, note that the contents of all lexical categories do not show any signs of becoming identical, as was the case when storage was not dependent on categorization in the simulation in Figure 25 above. In fact, as predicted, the dependence of storage on categorization encourages difference in the contents of distinct categories, as can be seen by comparing the exemplars of the lexical categories A and B at cycle 1000, shaded in gray above. At this point in this particular simulation, categories A and B are partially homophonous, sharing many identical exemplars. However, within a further 100 cycles, as evident in the cycle 1100 snapshot, we can see that the two categories have become distinct again.

On the other hand, visual inspection of the consensus of each lexical category at cycle 1500 suggests that the lexical items have still evolved to become more similar relative to their starting points, even in the context of selectional pressure in categorization to be contrastive. In fact, at cycle 1500, lexical items exhibit only four diphone sequences: *ka*, *ki*, *ik*, and *ak*. This constrained diphone set owes its form to the development of a large set of highly reliable correlations within this lexicon, three examples of which are provided below. These generalizations are written in the form of SPE-style rules, where [ ] represent segment slot boundaries.

i. [voi: 0] → [hi: .3] (voiceless segments are [k])

ii. [ ] → [voi: .1]/[voi: 0] ___ (segments after voiceless segments are voiced)

iii. [ ] → [voi: .1]/[hi: .3, voi: 0] ___ (segments after [k] are voiced)

Note that all correlations are operative within the grammar of the program, even those that contain other more general correlations. For example, although (i) and (ii) together imply (iii), within the program, these three generalizations contribute independently to target assembly. Recent evidence has shown that human speakers have access to similar ‘nested generalizations’ over their own lexicons in well-formedness judgments (e.g., Albright 2002).

Reliable patterns in the lexicon function to constrain target assembly to conform to those patterns, thereby reproducing themselves cycle after cycle. Generalizations (i—iii) above make the sequence [ka], for example, a basin of attraction, and act to influence similar sequences to surface as [ka]. For example, imagine in round 1501 of the simulation above that Performance modifies the speaker’s assembled target from lexical category D, *akak*, to *axak*. Imagine further that *axak* is nevertheless successfully recognized by the hearer as an example of the lexical category D, as it remains quite close to the majority of exemplars in that category. In the next round, the roles reverse, and the new speaker will produce a number of tokens of lexical category D. Let's imagine that one of those times the new speaker randomly chooses the *axak* exemplar stored in the previous round as the basis for target assembly for D. In this case, because there are so many sequences like [k] and [ka]in the lexicon that are close enough to [xa] to act as attractors, but so few actual [x] or [xa] sequences, there is nothing to compete with the [k] and [ka] attractors to anchor [xa] in place, such that the result of target assembly is very likely to be *akak*, rather than *axak*. The competition between basins of attraction explains why, for example [ka] rarely turns into [ki] in target assembly, even though [xa] will nearly always turn into [ka] in this lexicon. Although [ka] and [ki] are very similar and do exert a pull on one another, each is already sitting at the bottom of a deep basin of attraction formed by all the [ka] and [ki] sequences in the lexicon, respectively, so that each is only rarely turned into the other. Type frequency influences the depth of attractor basins in the program, so that as one sequence type becomes more widely distributed in the lexicon, it will have a proportionately stronger effect in influencing similar sequences. Hence, when the sequence [xa] happens to appear in the exemplar used as a basis for target assembly, even though there are a few [xa] sequences scattered about in different lexical categories due to errors introduced in Performance, their low frequency relative to [ka] sequences will mean that they have little chance to effectively compete in target assembly.

In Figure 29 below, I show graphical evidence that simulations including both cross-category influence in target assembly and selection for contrast evolve to exhibit constrained contrast. First however, let’s look at the results of another simulation, illustrated in Figure 27, that employs the conservative rather than the liberal procedure for lexical access.

Figure 27. Selection for contrast: conservative procedure

	Cycle 0	Cycle 500	Cycle 1000	Cycle 1500	Consensus at 1500
A	kaxa	xaka	xaka	xaka	
	kaxa	xaka	xaka	xaka	
	kaxa	kaka	xaka	xaka	
	kaxa	xaka	xaka	xaka	xaka
	kaxa	xaka	xaka	xaka	
	kaxa	xaka	xaka	xaka	
	kaxa	xaka	xaka	xaka	
	kaxa	xaka	xaka	xaka	
	kaxa	xaka	xaka	xaka	
B	xiki	xaka	kaxa	kaxa	
	xiki	kaka	kaxa	kaxa	
	xiki	kaka	kaxa	kaxa	
	xiki	kaka	kaxa	kaxa	kaxa
	xiki	kaka	kaka	kaxa	
	xiki	kaka	kaxa	kaxa	
	xiki	kaxa	kaxa	kaxa	
	xiki	kaka	kaxa	kaxa	
	xiki	xaka	kaxa	kaxa	
C	ixax	axax	axak	axak	
	ixax	axax	akak	axax	
	ixax	axax	axax	axak	
	ixax	axax	axax	axak	axak
	ixax	axax	axax	axak	
	ixax	axax	axak	axak	
	ixax	axax	axix	axak	
	ixax	axax	akak	axak	
	ixax	axax	akak	axak	
D	akik	akak	akak	kxak	
	akik	akak	kkak	kxak	
	akik	akak	axax	kxak	
	akik	akak	axak	kxak	kxak
	akik	akak	axax	kxak	
	akik	akak	akax	xxak	
	akik	akak	kkax	kxak	
	akik	akak	akak	kxak	
	akik	akak	xkak	kxak	

As we saw in Figure 26 above, when cross-category influence on target assembly and selection for contrast in storage can interact, the lexicon fails to devolve to uniformity. In fact, as did the liberal procedure for categorization, the conservative process can be seen to actively militate against encroaching homophony: as can be seen in the gray box, at cycle 1000 lexical categories C and D have become close enough that many of their exemplars cannot be reliably categorized. However, because greater contrast leads to more reliable categorization, and because the probability of consistent storage is higher when categorization is reliable, any more contrastive exemplars that happen to arise under these conditions are likely to spread (in this case, such as the few *k*-initial D exemplars). At 1500 cycles, we find a set of lexical categories in which contrast has been maintained despite some significant drift from the starting points at cycle 0. As before, however, visual inspection of the resulting forms suggests that they share more sequences than the forms in the starting lexicon. Again, this increase in similarity across the lexicon over many cycles suggests that selection for contrast in categorization on the part of the hearer has not eliminated all effects of cross-category influence on speaker outputs, but has simply prevented them from eventually bleeding all contrast from the lexicon.

In Figure 29 below, I present a graph of this increase in similarity across a lexicon over the course of a simulation. The simulations for the figure were carried out with a six-entry lexicon that at the start contained fully randomized exemplars, an example of which is shown in Figure 28, for 500 cycles. The lexicons were randomized at the beginning to avoid any inadvertent, consistent bias toward a particular outcome.

Figure 28. Sample randomized starting lexicon

A	B	C	D	E	F
ixxx	kxka	xaaa	xxkx	axxi	kxxk
xaki	iiki	kxix	aakx	aiki	aixk
aiax	kkxi	iaax	xkkk	aaia	kaii
xaik	kiak	akxa	aakx	xxaa	kkxa
kaxx	xiii	kiax	xkai	ixik	xiii
kaax	iiax	akax	xaka	xxak	akxi
iakx	kiii	xaax	aika	iiax	kikx
iikx	aaka	kika	iikx	xxax	xaii
akia	xxia	akaa	kxka	kaxi	aaii

To provide a measure of similarity across lexical entries at each cycle, the relative entropy of the system (Shannon 1949) was calculated. The entropy of the system is a measure of the degree of predictability, or redundancy, in the system. The true entropy of the system at a given point was approximated by calculating the entropy on the basis of two-segment sequence (‘diphone’) frequencies. When we asses the predictability within the system on the basis of diphone frequencies, we measure predictability not only in terms of the frequencies of individual segments, but also of segment pairs. The calculated entropy is reported in Figure 29 below in the form of relative entropy by dividing by the calculated maximum entropy of the system[19]. The maximum relative entropy, representing a system in which every segment sequence is equiprobable, is therefore 1. The minimum entropy, representing maximum predictability, is not zero for this system because the symbol sequences are constrained by the simulation to follow the order LeftEdge, Segments…RightEdge, with the result that every sequence must contain at least three different symbols. As it turns out, the minimum entropy for a four segment word, composed from a pool four segment types plus two distinct edge markers, is 0.25.

Figure 29 shows the results of five distinct simulations simultaneously, all starting from randomized six-lexical category lexicons like that shown in Figure 28 above, but differing in the presence or absence of cross-category

---

[19] The maximum possible entropy of a system of elements is given by the logarithm in base 2 of the number of kinds of elements. In this case, the symbol set comprises four segments and a left and right edge, for a total of six elements.

influence in target assembly and selection for contrast in categorization. The conditions for the simulations illustrated in Figure 29 are given below:

1. No cross-category influence in target assembly, no selection for contrast.

2. Selection for contrast (conservative procedure) in categorization, no cross-category influence in target assembly.

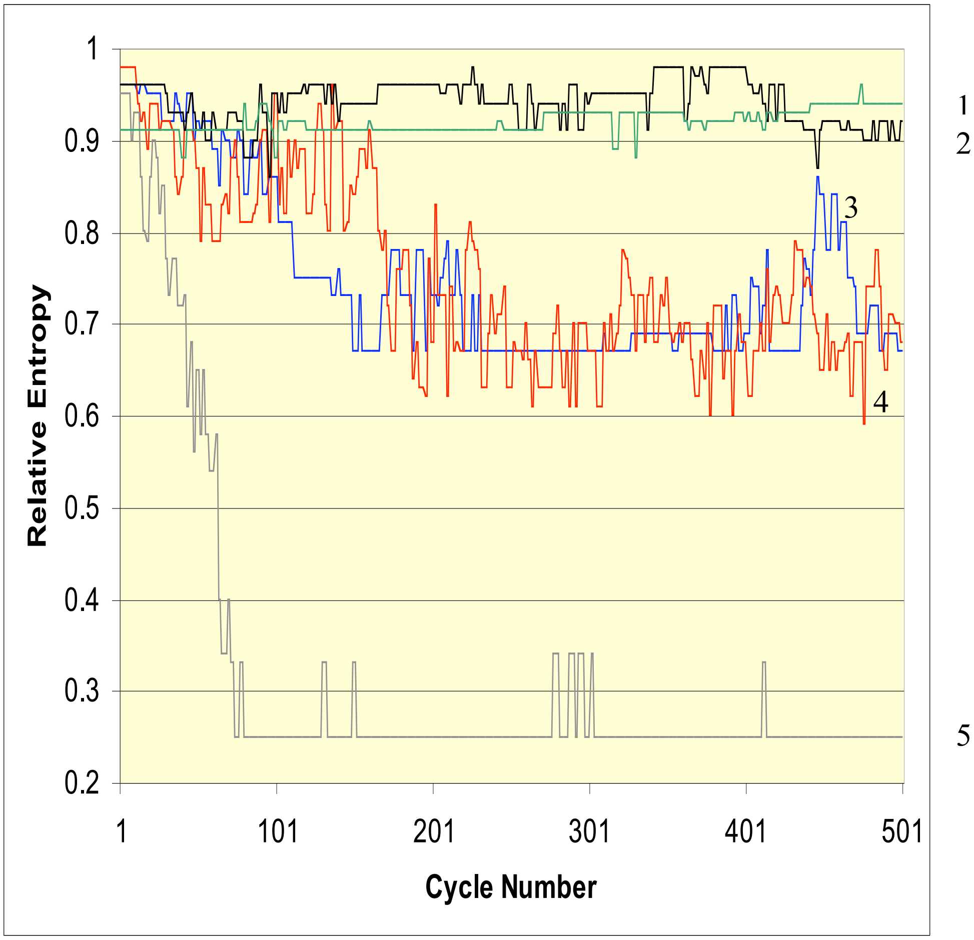
3. Both cross cross-category influence in target assembly, and selection for contrast (conservative procedure).

4. Both cross cross-category influence in target assembly, and selection for contrast (liberal procedure).

5. Cross cross-category influence in target assembly, but no selection for contrast.

In both conditions (1) and (2) we can see in Figure 29 that the entropy remains near one, indicating that the degree of predictability within the lexicon is near the minimum possible. This is not unexpected, since in condition (1) each lexical category evolves completely independently, and in condition (2) the only constraint on lexical category evolution is that they not come too close within the relatively large contrast space provided. In contrast, in condition (5) where cross-category influence encourages targets to resemble patterns already present in the lexicon, but contrast plays no role in categorization, all lexical categories rapidly evolve to be composed of one segment, in this particular case [a]. The relative entropy under these conditions approaches the minimum possible given fixed word-edge elements, .25. Random changes are still being steadily introduced in Performance, as can be seen in the occasional small spikes in relative entropy, but the strong pattern-consistency within the lexicon remains a deep basin of attraction, with the result that these random alterations have little chance of becoming established.

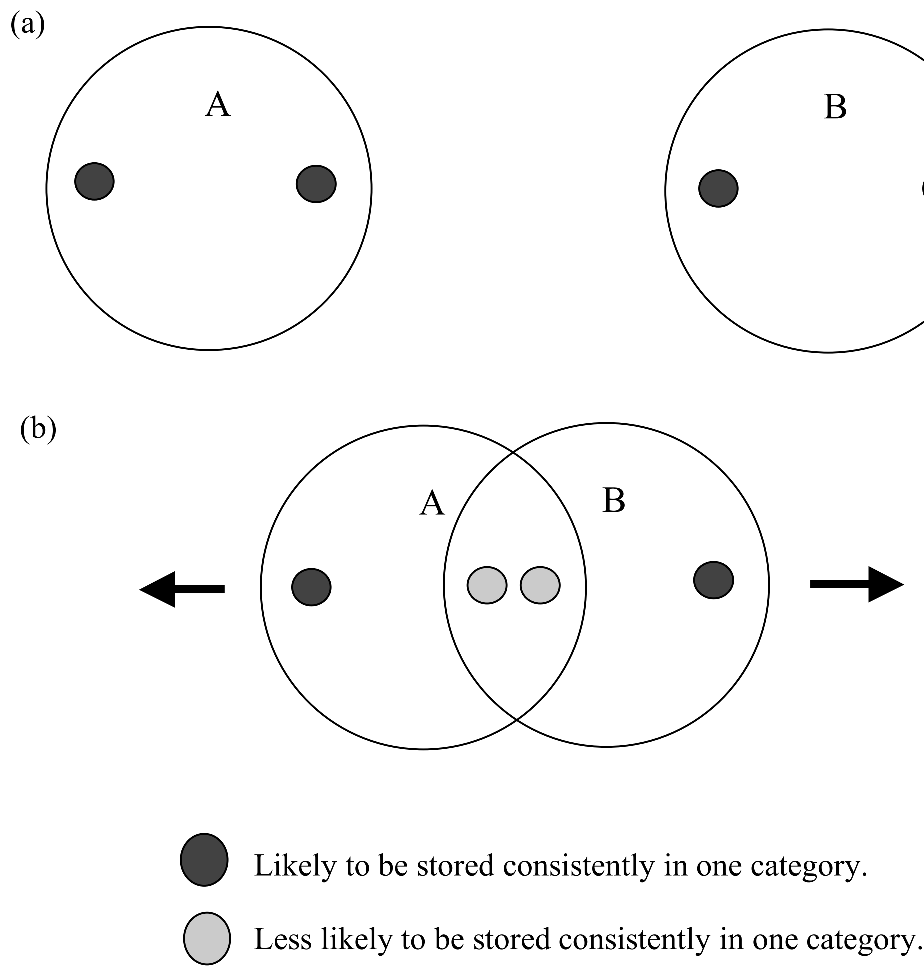
Figure 29. Diphone entropy is influenced by both pattern reinforcement and selection for contrast.



In conditions (3) and (4), just as in (5) cross-category influence in target assembly begins to bring categories closer together from a starting relative entropy around .9, but the fact that less contrastive speaker-outputs are less consistently categorized, results in abias in categorization towards more contrastive outputs. The result is that the slide toward uniformity is interrupted, producing an equilibrium state in which lexical categories orbit about one another within the contrast space, staying close, but rarely overlapping completely. Both categorization procedures, conservative and liberal, provide similar results. However, one difference between them that can be seen in Figure 29 relates to the different relative stability of the lexicon under the two procedures. Because more extreme variants of a category are less likely to be categorized and stored under the conservative procedure for categorization, the lexicon exhibits considerably more hysteresis (i.e., dependence on prior states) than under the liberal categorization procedure. The resulting greater stability in lexical categories can be seen in the smoother changes in relative entropy in simulation (3) relative to (4) in Figure 29. A related difference in behavior has to do with the rapidity with which an evolving lexicon can change states within the simulation. Because the rate of sampling different sequences within the available contrast space is lower under the conservative procedure than the liberal procedure, it takes longer for an evolving lexicon to 'find' and move between distinct local minima. For example, in different runs of the simulation types represented in Figure 29, lexicons under the liberal procedure generally move from the starting high-entropy state to some lower state within a few hundred cycles. Under the conservative procedure however, the point at which a lower entropy state is discovered varies over a much larger range, from a few hundred cycles (as in the example shown above) to several thousand.

The results presented in the figures above suggest that the conflict between pressures toward uniformity in production on the one hand, and pressures to maintain functional contrast in the lexicon on the other causes a cyclically updated lexicon to evolve to exhibit 'constrained contrast'. As can be seen in Figure 29, optimization of the conflict results in lexical categories packing relatively close to one another, but not so close that categorization efficiency falls too low. This equilibrium is reached because the closer categories get in sequence space, the higher the storage consistency of exemplars that are further from the category boundary relative to the storage consistency of exemplars closer to that boundary, as illustrated below in Figure 30. Due to this growing asymmetry in which forms get stored as categories get closer, their center-points tend to be driven back apart. The resulting compromise is a lexicon that reuses as many diphone sequences as possible, while still supporting contrast between lexical categories.

Figure 30. Overlapping categories are driven apart by asymmetric exemplar storage patterns.



Large circles represent the boundaries of exemplar clouds belonging to categories A and B, positioned within some contrast space. Small black and gray circles represent exemplars. In (a), the two clouds are distant enough in contrast space that each depicted form is highly likely to be re-stored in its respective category if it is transmitted to the other speaker and then returned in a subsequent round. In (b) however, categories A and B have gotten close enough to one another that storage consistency is no longer the same for the depicted exemplars. For example, if a gray exemplar in A is passed to the other speaker in one round and then returned in the next, it has nearly as great a chance of being categorized as an example of category B as it does as an example of category A. The black exemplar from A on the other hand is much more likely to be correctly categorized. Because exemplars are being randomly lost from a category every time a new exemplar is stored, the exemplar types that most often contribute to a given category will come to dominate that category, effectively moving the center of the category further away from other competing categories.

2.4.3 Comparison to previous work modeling contrast maintenance

As far as I know, no simulations of patterns in linguistic systems to date have directly modeled the interaction between successful reference to external categories and patterns in contrast as carried out here. Instead, the degree of contrast has been either predefined within the model, or monitored within a simulation and directly provided to simulated participants. As an example of the former, Lindblom (1986, 1992) used experimentally derived values for 'discriminability' (Plomp 1970, Bladon and Lindblom 1981) and model-derived values for 'articulatory cost' (Lindblom et al. 1983) to arrive at an equation for optimizing these two factors. This equation was used to find the best set of a fixed number of syllables out of a predefined, larger set. The finding that syllable sets identified by this algorithm corresponded closely to crosslinguistically common contrasts is consistent with the proposal that discriminability and articulatory cost interact to influence inventories, with something close to the relative weights used in the equation.

Redford et al's (2001) simulations demonstrating emergence of optimal syllable types proceed as well through an external evaluation of the fitness of words on the basis of distinctness. In these simulations, the degree to which competing vocabularies contributed to a next generation of vocabularies was determined within the simulation algorithm in part by the number of homophonous items they contained.

As a further example of direct provision of contrast data within a simulation, in de Boer's (2000) simulations of the development of vowel inventories, two parallel channels of communication are provided between agents. Using the articulatory/perceptual channel, a speaker produces a vowel from a category, which a hearer then matches to one of its categories. Using a second, 'non-verbal' channel, the speaker and hearer compare the source and match categories; if they are the same, the communication has been successful, if not, a failure. Success or failure result in different updates to the hearer's vowel system, which then over time result in the preferential development of certain patterns in vowel contrast over others.

Likewise, simulations of the evolution of form-meaning structure involve either two parallel channels, one of which serves as a means of direct feedback on meaning (Steels 1997), or involve communication about a 'scene' that both speaker and hearer can observe (e.g., Hurford 1989, Kirby 2000, Batali 2002). In the latter, the comparison of the directly observed scene and the communication describing the scene provides the hearers within the simulation the feedback necessary to direct modifications of their own form-meaning structures.

Janet Pierrehumbert's simulations of the interaction of exemplar-based phonetic categories have shown that exemplar theory can account for a number of observations concerning the influence of low-level phonetic detail on category contents, and on the interaction of categories in lenition (2001, 2002). In her simulations, however, no mechanism acts directly or indirectly to preserve contrast at all, with the result that a single category eventually absorbs all others.

If we assume that the notion of contrast only has substance in the context of an actual form-meaning pairing, then this is of course reasonable, because sub-morphemic categories do not have independent meaning of their own, but only contribute to marking meaning difference in larger sound sequences. In the simulations presented here, distinctions between feature categories are preserved, but only indirectly through the preservation of difference between lexical categories. Maintenance of difference between lexical categories, in turn, is driven not by any direct measure of contrast in the system, but only indirectly through differing patterns of storage of contrastive, versus less contrastive forms, as illustrated above in Figure 28.

Within linguistics, the notion that contrast maintenance is an indirect effect of contrast's effect on a hearer/acquirer's categorization behavior has previously proposed by Gregory Guy (1996) on the basis of corpus data.

Examining various deletion processes in English and Brazilian Portuguese production data, Guy finds *no* convincing evidence that speakers systematically avoid deletion of segments or features just in case a morpheme would be rendered unrecoverable. How then to explain the many grammatical processes evident in language that appear to function to preserve morphological contrasts (see e.g., Kiparsky 1982, p. 91ff for examples)?

Guy cleverly begins his account by noting that data from production corpora will always underestimate the true extent of speakers' failure to produce a given meaningful contrast, because when a contrast is completely lost in an utterance, the transcriber has no way to reconstruct and recover this loss. For example, if a transcriber perceives the utterance 'I cook the chicken', s/he is likely to simply transcribe it as such, even if the speaker actually intended the sentence to be in past tense, but elided the [-t] past tense marker. Guy notes that language acquirers are no different from transcribers, such that the perception data from which a language learner develops a grammar will be biased towards the more contrastive utterances in the production data set. This steady selection of more contrastive forms in the categorized utterance set upon which acquisition is based should result in a tendency for grammatical processes to emerge that appear to function to preserve contrast, when they in fact only act to reproduce the patterns in the data set that the acquirer perceives. Guy provides an example from Mecklenburg German, taken from Kiparsky (1982), in which final unstressed schwa can be deleted from nouns, except when the schwa is the sole marker of plurality:

i. [gast] ~ [gEst↔] $\rightarrow$ [gast] ~ [gEst] ‘guest(s)’

but:

ii. [spEr] ~ [spEr↔] $\not\rightarrow$ [spEr] ~ [spEr] ‘javelin(s)’

Guy proposes that this state of affairs may come about if a child acquiring Mecklenburg German is likely to miscategorize a plural form with a deleted final schwa as the singular form, thereby failing to notice the deletion at all. The child then develops a grammar corresponding to the perceived input: final schwas are often deleted, but tend to be preserved in those forms in which the final schwa is the only marker of plurality. This tendency promotes development of a community of language users in which schwa deletion appears to be sensitive to the functional redundancy of the plural marker.

Under Guy’s proposal, the grammar does not develop through direct reference to contrast, but is only indirectly influenced through contrast’s influence on categorization patterns. This indirect means of maintaining contrast is similar to the mechanism of contrast-maintenance operating within the simulations presented in this chapter. In this model, the simulated speaker does not filter outputs with respect to contrastiveness, but only with respect to patterns that exist within its own lexicon. Contrast becomes relevant only at the point of categorization by the hearer. Highly contrastive forms are stored consistently in one category, thereby concentrating their influence on the evolution of the lexicon in that category. Less contrastive forms, on the other hand, dilute their influence among multiple categories, and therefore have less effect on the evolution of any single category. The result is indirect selection for contrast.

This brings us to the first of the general questions about constrained contrast in phonological systems posed at the beginning of this section above -- why do individual phonological systems tend to use idiosyncratically defined sub-regions of the total available contrast space? Given that many phonological systems have velar fricatives, voiced bilabial stops, or high front round vowels, why don't they all have them? Over the last century, explanation for the tendency for languages to make use of constrained subsets of the possible contrasts has generally been based in the notion of an innate, universal grammar whose application naturally tends to categorically limit the number of features active in a language (e.g., Jakobson 1952, Halle 1959, Prince and Smolensky 1993, Kenstowicz 1994, pp 57-61). However, in support of previous work, in particular that of Lindblom (1992), Bybee (2001) and Pierrehumbert (2001a), the results presented here suggest that pattern reinforcement in language use and transmission should also function to allow the emergence of idiosyncratic asymmetries across languages. In the simulations presented above and in Chapter 1, we've seen that cross-category influence on production results in a tendency for a system to evolve to employ a limited range of the elements available -- even when there is nothing intrinsically better or easier about one subset of the available elements relative to any other. Here, the choice of subset is essentially random: myriad random events initiate self-reinforcing basins of attraction that then shape the pathway a simulation takes. Eventually, some segments and segment sequences are lost, and the use of others is expanded.

The contrast-limiting properties of the model are reminiscent of natural phonological systems in which contrasts tend to be drawn from more or less contiguous regions of the abstract space describing possible human phonological contrasts, rather than from widely scattered points in that space. In the following section, I go further to show that when we incorporate the notion that some parts of that space are less congenial to language use in some way than others, we find that evolving lexicons avoid these regions when possible, exploit them only when contrast would be otherwise difficult to maintain, and if they must exploit them, then they do so to a lesser degree than contrasts in the more congenial regions.

2.5 Contrast-patterns and Markedness

In all simulations shown up to this point, changes in Performance have been random and context-free, targeting every feature equally. In the following section, the degree of random, context-free alteration will remain the same, but in addition, two context-sensitive biases will be imposed, making some pathways of change more likely than others. The first bias, which we'll call NoCoda for convenience, increases the probability of change of word-final consonants relative to word-final vowels, and the second, which we'll call Onset, increases the probability of change of word-initial vowels relative to word-initial consonants[20]. Both biases are set at 10%, meaning that a segment

---

[20] Arguments have been made that the preponderance of syllables with onsets over those with codas in the languages of the world may not derive directly from a bias in favor of onsets (cf Bybee 2001, pp 206-9). I use these metaphorical 'biases' here as tests because they are symmetric in the word, and because within the feature system used they do not produce patterns sharing features in common. Hence, these biases enjoy equal and independent

in an assembled target in violation of one of these biases will be altered randomly 10% of the time. Because in this section we'll be exploring what happens when the contrast space gets tight, we will use words of three, rather than four segments in order to reduce the contrast space available to the lexicon.

Ceteris paribus, the less stable a property is in transmission, the less likely it is to persist in an evolving system. Hence, we expect that over time, these biases should result in lexicons that favor consonant-initial, and vowel-final forms. In a first example simulation, we see how the lexicon evolves when there is plenty of contrast space for lexical categories to be differentiated . In Figure 31a below, the consensus lexical categories of a lexicon containing 6 categories are shown at 0, 400 and 800 rounds in a simulation in which both the NoCoda and Onset biases are in place. Out of 64 total possible 3-segment words drawn from the segment inventory of [k x i a], there are 16 possible words that violate neither bias, that is, both begin with [k] or [x], and end with [i] or [a]. A lexicon with six lexical categories therefore might be expected to be able to find six three segment sequences that are sufficiently different from one another to support efficient

---

opportunities to produce patterns that become entrenched within the evolving categorization. And in fact, that is what we find. When a simulation starts with a seed lexicon containing six lexical entries with a distribution of bias-violating sequences, the system rapidly evolves to a state in which neither of these biases are violated by the consensus contents of any lexical category. This occurs because the contrast space is large enough that each one of the bias-violating lexical items can easily be changed into something non-bias-violating without bumping into some other category in the process[21].

Figure 31a. Influence of NoCoda and Onset biases: consensus forms

Lexical category	Cycle 0 6 bias violations	Cycle 400 4 bias violations	Cycle 800 0 bias violations
A	kaa	ka<u>k</u>	kaa
B	xii	xxi	xii
C	xk<u>x</u>	xki	xxa
D	<u>i</u>ki	kki	kki
E	<u>a</u>x<u>x</u>	<u>a</u>x<u>x</u>	kxi
F	<u>i</u>a<u>k</u>	xa<u>k</u>	xai

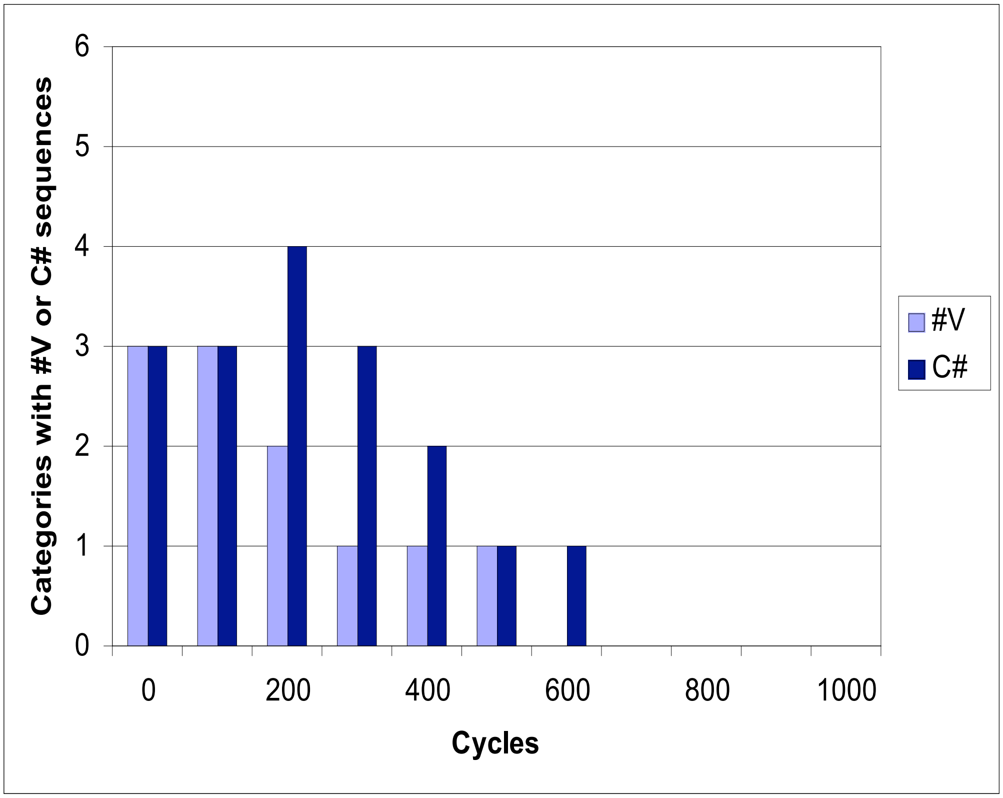
(Underlined segments violate Onset or NoCoda)

lexicon.

[21] Each simulation type presented in this chapter was carried out many times from a variety of starting conditions; all results shown represent statistically significant, typical equilibrium points.

Figure 31b shows a summary of the same simulation in the form of the number of violations of each type in the consensus lexical entries in the lexicon at each 100 cycles.

Figure 31b. Influence of NoCoda and Onset biases: relative frequencies of forms



Biases in Performance increase the rate of change *out* of regions of the available contrast space corresponding to the sequences [left word-edge, vowel], and [consonant, right word-edge], with the result that lexical entries tend to evolve away from these sequences. At a point where most lexical entries have done so, however, the absence of these sequences in the lexicon becomes a reliable pattern, with the result that it then becomes more difficult to evolve *into* these regions as well. Imagine, for example, that a lexicon largely conforms to Onset, meaning that nearly every [left word-edge] in the lexicon is followed by [voi: 0] in the word-initial segment. If an exemplar is chosen for target assembly in which the initial segment has the voicing feature value [voi: .1], the value of that segment is liable to be changed to [voi: 0], 'repairing' the incipient violation of Onset. In other words, once a pattern becomes sufficiently reliable, by whatever means, it becomes 'productive', continuing to reinforce that pattern in target assembly.

2.5.1 Productivity in the simulation

The model for the simulations in this dissertation, while non-deterministic, is also generative in the sense that patterns can be productively extended to novel forms. The productivity of patterns in the simulated lexicon is illustrated below in a simulation with an eight-category lexicon that had already evolved to reliably conform to the bias Onset, meaning that the majority of exemplars in every lexical category began with consonants. At the point when all lexical categories had come to satisfy Onset in most of their exemplars, the simulation was paused, and one token of a new lexical item, [ixa] was introduced to both speakers – imagine, for example, that this new word was pronounced once for them by a third party – upon which both speakers stored that perceived pronunciation as a *single* exemplar in a new lexical category, 'J'. The simulation was resumed, whereupon both speakers began pronouncing this new word for each other as for every other lexical category in their lexicons. For the first four cycles after the introduction of a new word, before either speaker had built up a sufficient set of exemplars in the new lexical category to allow efficient identification and categorization, the program allowed the two speakers to positively identify their utterances of the new word for each other as an exemplar of the intended category, in this case 'J'.

In this particular simulation, the new word [ixa] begins with a vowel, violating the pattern previously established in the lexicon. To make the subsequent behavior of the speakers more informative for our purposes, at the point that the simulation was resumed, Performance was *entirely deactivated* to allow us to attribute any change in the form of the new word solely to alterations introduced in target assembly. Figure 32 below shows the contents of Speaker 2's category 'J' at the beginning of each cycle over the course of the first 12 subsequent cycles (every other cycle is shown because a given speaker’s lexicon is only updated every 2 rounds as they serve as ‘hearer’).

Figure 32. Productivity in the simulation, given a lexicon with no #V sequences.

Cycle Number	0	2	4	6	8	10	12
Exemplars in Speaker 2’s Lexical Category ‘J’	ixa	ixa	ixa	ixa	ixa	xxa	kxa
		xxa	xxa	xxa	xxa	kxa	kxa
		kxa	kxa	xxa	xka	kxa	kxa
			kxa	kxa	kxa	kxa	kxa
			kxa	kxa	kxa	kxa	kxa
				kxa	kxa	kxa	kxa

/ixa/ → [kxa]

Both speakers’ category ‘J’ began the re-started simulation at Cycle 0 with a single exemplar, [ixa]. In cycle 0, Speaker 1 produced two utterances corresponding to lexical category ‘J’, and, in this initial phase, indicated directly to Speaker 2 that these utterances were members of category ‘J’, who stored them as such. As can be seen in Speaker 2’s lexicon at the beginning of cycle 2, it is clear that neither of Speaker 1’s productions of this word began with the original vowel, but rather substituted an initial [k] or [x]. This occurred because nearly all the stored exemplars in Speaker 1’s lexicon began with consonants, with the result that the initial segment sequences of the new word [ixa] could not serve as a sufficiently reliable pattern to countermand the more reliable consonant-initial pattern in the lexicon. By cycle 4, the program required both Speaker 1 and 2 to rely on categorization rather than 'deixis' for identification and storage, but by this time there were already sufficient consonant-initial exemplars of category 'J' to allow efficient categorization and storage of consonant-initial variants. In this particular example, the initial [ixa] eventually settled down to the form [kxa].

By comparison, in a parallel simulation, identical except that two of the eight original lexical categories did contain vowel-initial forms, the same newly introduced word [ixa] survived with its initial vowel intact. This is not unexpected, because even though there were more consonant-initial forms in this lexicon than vowel-initial forms, this pattern was not robust enough to significantly intervene in target assembly.

The important point to take from the simulation illustrated in Figure 32 is that there are two functionally linked, but mechanistically independent sources of patterning in this program: biases in Performance, and the reinforcement and extension of extant lexical patterns through cross-category influence in target assembly. Because the latter is agnostic about what patterns are reproduced, all else being equal, the lexicon evolves such that cross-category influence eventually comes to recapitulate exactly those biases in

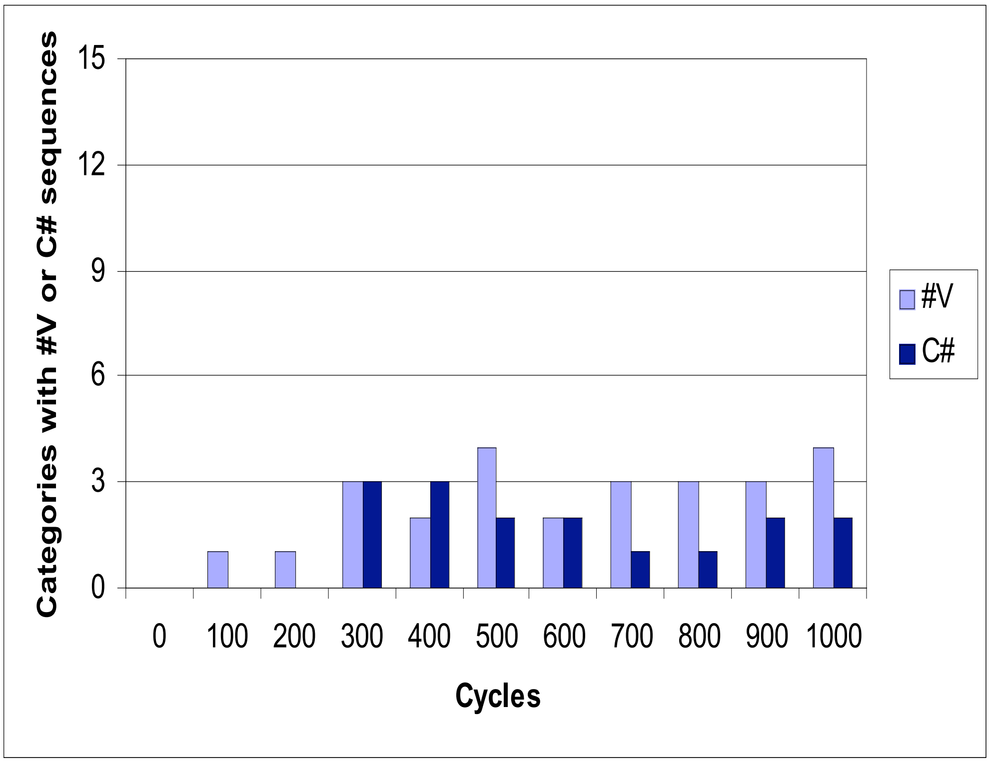
Performance which successfully initiated asymmetries in the lexicon in the first place.

2.6 Typologies and Markedness

Above, we saw that when certain regions of the available contrast space are less hospitable than others, the lexicon tends to evolve away from those regions. However, we might expect that the ability of the lexicon to eschew the marked regions of contrast space altogether would be predicated on contrast remaining high enough for categorization to remain efficient. What happens, then, if we increase the number of lexical items, but leave the size of the available contrast space constant?

Figure 33 shows the results of a typical simulation similar to that above, but in which the number of categories in the speakers' lexicons has been increased from 6 to 15, and in which pattern reinforcement has been disabled, in order to illustrate the patterns that develop just through the action of biases in Performance, and selection for contrast through category competition. In addition, for the purpose of illustration, the lexicons are seeded with 15 out of the 16 possible C-initial, V-final 3-segment forms.

Figure 33. Evolution of a 15 category lexicon *without* pattern reinforcement.



The results of the simulation illustrated in Figure 33 differ most strikingly from those of Figure 32 in that rather than being avoided, the less-fit V-initial and C-final forms are innovated, and remain consistently part of the speakers' lexicons. This occurs because, given the degree of mismatch allowed in category matching (see Figure 23 in section 2.3.6 above), 15 categories is too tight a fit for the 16 possible fit slots in the available contrast space, and so contrast selection begins to favor exploitation of the less fit initial vowel and final consonant sequences. There are two additional important points to take from this graph. First, although less-fit sequences are exploited in the service of increasing contrast, they remain in the minority, with an average of about 20% of the categories using any given less fit sequence[22]. Second, both kinds of less fit sequences, initial vowels and final consonants, are innovated and exploited. We'll see below that this latter pattern changes when pattern reinforcement is reintroduced.

As we've seen above, Performance biases discourage lexical items from exploiting marked regions of the contrast space both directly and indirectly. First, given a lexical item that already contains marked sequences, biases in Performance increase the rate of feature-value changes away from these sequences relative to unmarked sequences. Secondly, to the extent that a marked sequence is rare in the lexicon, pattern reinforcement in target assembly will make it more difficult for any assembled target to surface containing that marked sequence. Biases in Performance therefore encourage lexical categories to evolve *away* from marked regions of contrast space,

[22] In simulations starting from the 'opposite' position, in which every lexical category is seeded with vowel-initial and consonant-final exemplars, the same pattern is reached, with a low level of exploitation of these sequences.

while bias-derived patterns in the lexicon discourage lexical entries from evolving *into* previously unexploited regions of contrast space. However, pressure from pattern reinforcement on a target *not* to be assembled with a marked sequence diminishes rapidly if that sequence becomes more established in the lexicon. Therefore, the initial establishment of a marked sequence in the lexicon should be difficult, but once established, it should become progressively easier to extend that sequence to other lexical entries.

Imagine now a lexicon that conforms to both of the two biases in Performance, Onset and NoCoda, with the result that both of the correlations below are robust. Written in SPE style, we have:

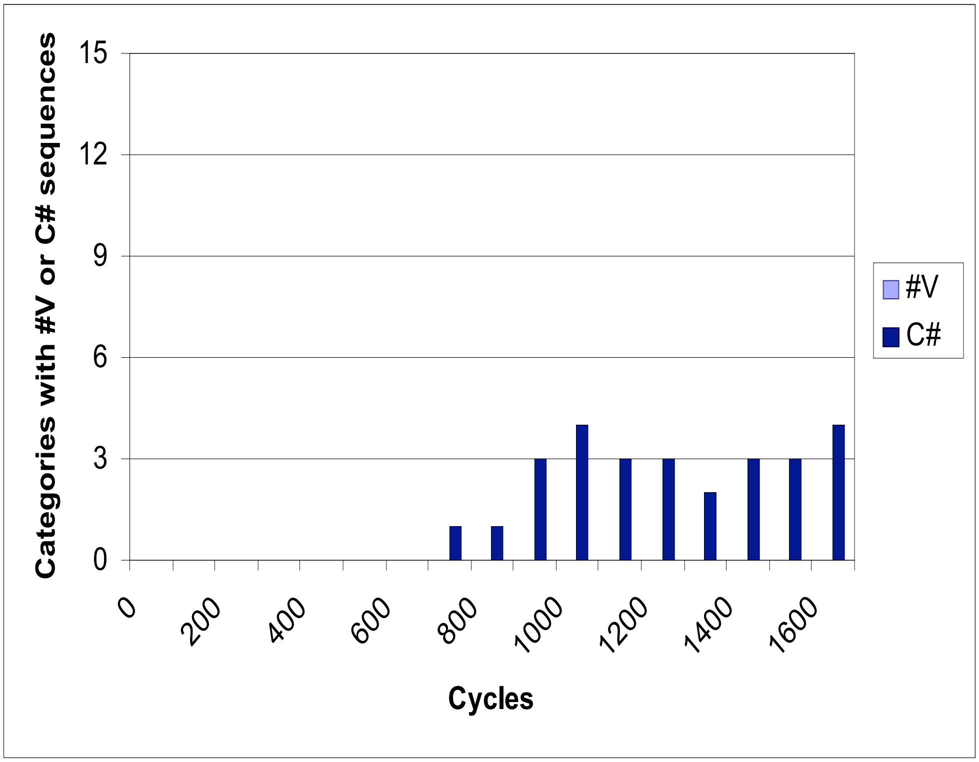
Onset: [ ] → [voi: 0]/[Left Word-edge] ___

NoCoda: [ ] → [voi: 0.1]/___ [Right Word-edge]

Imagine further that the available unmarked regions of contrast space are cramped, such that categorization of incoming speaker-outputs is sufficiently inefficient that lexical categories are under more pressure to drift further apart from selection for contrast, than they are under pressure to drift together from cross-category influence. Under these conditions, we would expect that the occasionally stored speaker-output emerging from Performance with an initial vowel or final consonant should be able to serve as the nucleus for a shift in the center of a lexical category into a marked region of contrast space. However, when that stored vowel-initial or consonant final form has its turn as the basis for target assembly, as we saw above, the robust pattern in the rest of the lexicon makes it unlikely that this more contrastive sequence will actually appear in the assembled target. Hence, our imaginary lexicon may continue on for quite some time in a state of insufficient contrast, simply because it cannot break into new regions of the available contrast space. However, the randomness inherent in Performance, storage and target assembly make it likely that eventually a few Onset- or NoCoda-violating exemplars will happen to collect in a given lexical category. At this crucial point, the filter against sequences violating that bias in target assembly begins to weaken, allowing progressively more bias-violating sequences to surface in assembled targets, and thus what began as a tiny, random leak into a marked region of contrast space, becomes if not a torrent, at least a respectable flow. Crucially, because Onset and NoCoda produce distinct correlations in the lexicon, when one correlation becomes unreliable in the lexicon, only the corresponding marked region of contrast space becomes neutral in terms of target assembly. Hence, when our imaginary lexicon breaks out of the cramped, unmarked region of the available contrast space, it is likely to do so by innovating either vowel-initial words, or consonant-final words, but not both. However, note that the biases against such sequences remain operative in Performance, so even at a point when target assembly no longer discriminates against a given Performance bias-violating sequence, allowing it to spread in the lexicon, we expect that sequence to remain relatively rarer in the lexicon than less marked sequences. Nonetheless, we expect the progressive slackening of pattern-derived inhibition against using a marked sequence, as it becomes more frequent, to result in the following tendency: if a lexicon makes use of a marked segment at all, it is likely to use it relatively frequently, or put conversely, a lexicon is unlikely to contain just a few examples of a given marked sequence.

Figure 34 below illustrates a typical simulation identical to that presented above in Figure 33, except that pattern reinforcement has been reintroduced.

Figure 34. Evolution of a 15 category lexicon *with* pattern reinforcement.



There are two points to take from the above figure. First, only one less-fit sequence-type is innovated in this simulation, rather than both; this is a statistically significant deviation (see discussion of Figure 35 below). Second, the first innovation of a less-fit sequence occurs much later (~ 800 cycles) than in the previous simulation (Figure 33; ~ 100 cycles). Both of these observations bear out predictions made in the discussion above:

- Pattern reinforcement should favor targeted exploitation over more diffuse exploitation of less-fit regions of contrast space.
- Pattern reinforcement renders the initial alteration of an entrenched pattern more difficult.

In Figures 35 and 36 below, I show a summary of the results of two hundred independent simulations run under identical conditions that collectively illustrate this phenomenon. Each simulation follows the evolution of a lexicon containing 15, initially randomized, three-segment lexical items each drawn from the set [k x i a] for 2000 cycles. Performance in the simulations administers both Onset and NoCoda biases; all simulations use the liberal categorization procedure. The question here will be whether these lexicons that have bias-violating forms distribute their violations randomly between sequence types, or whether they instead tend to concentrate their violations in one region of contrast space, as predicted in the discussion above.

Taken together, the 200 simulations ended with 262 total categories with consensus #V sequences, and 252 total categories with consensus C# sequences, out of 3000 categories possible (200 lexicons X 15 lexical categories). Given the null hypothesis that the distribution of each sequence type was random, i.e. binomial, the expected proportion of lexicons with one, neither, or both sequence types was calculated and compared to the observed distribution. For example, the null hypothesis of a random distribution of sequence types across lexical categories predicts that there should be ~50 lexicons containing no categories with #V sequences, and ~50 lexicons containing no categories with C# sequences. Instead, we find 85 and 90 lexicons with no categories containing #V or C# sequences, respectively.

Figure 35. Distribution of contrast-types across 200 identical simulations with pattern reinforcement and selection for contrast.

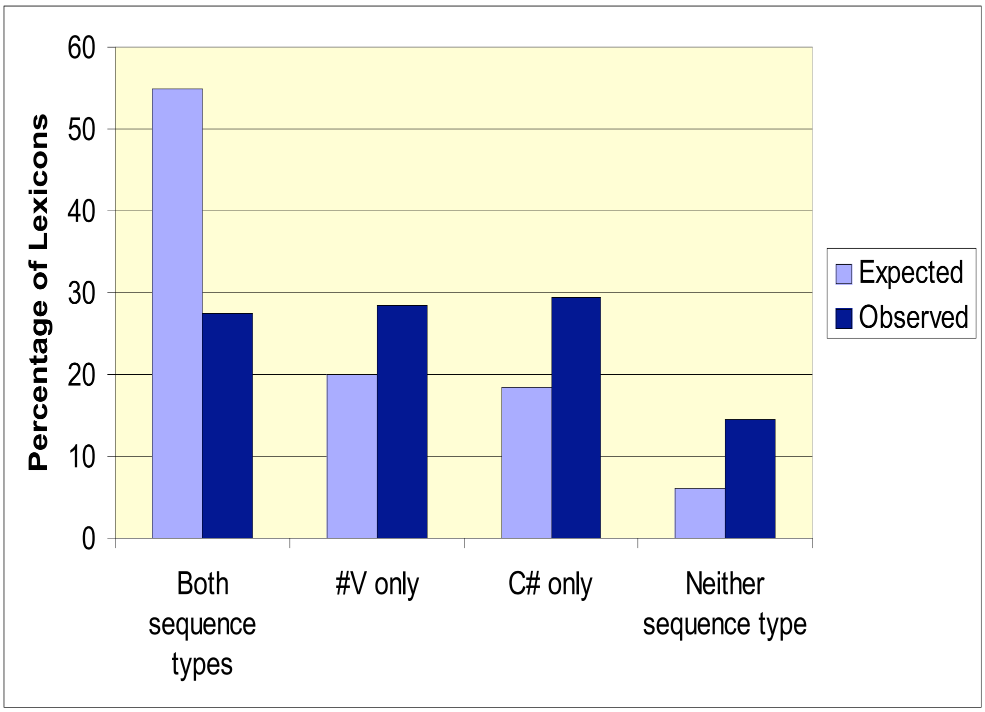
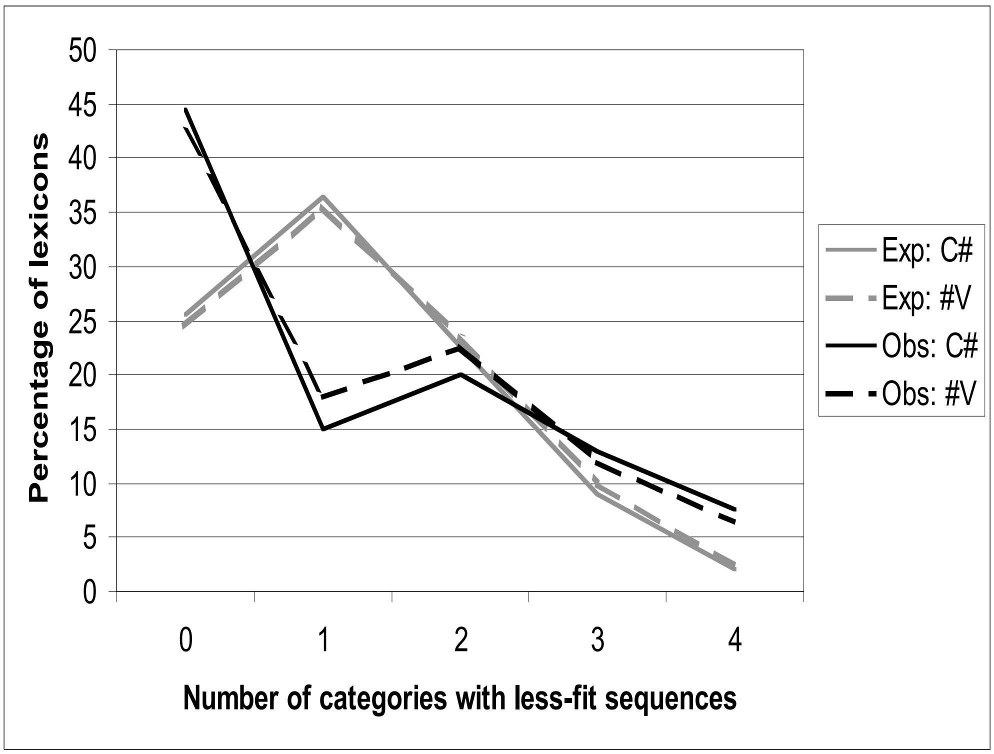


Figure 33 shows that the proportion of lexicons containing one or the other less-fit sequence is higher, at the expense of those lexicons containing both marked sequences. This tendency towards categorical exploitation of less-fit contrasts is significant (Chi^2: $p \ll .001$).

Figure 36 illustrates the same data in a slightly different way, showing the expected, versus observed, proportion of lexicons containing zero, one, two, three or four categories containing a less-fit sequence.

Figure 36. Distribution of less-fit sequences across lexicons



Inspection of Figure 36 shows that the observed distribution is skewed away from the expected binomial distribution towards categoricity, in that lexicons show an increased tendency to either have no sequences of a given less-fit type, or multiple sequences of that type. The double maxima in the observed distributions suggests, as we reasoned in the paragraphs above, that cross-category influence in target assembly would promote an all-or-nothing pattern in exploitation of less-fit regions of contrast space. Recall from the discussion above that the initial usage of a novel sequence should be strongly discouraged by cross-category influence in target assembly, but that once established, it should be able to spread more easily to other lexical items in the lexicon. This 'first step is the hardest' property should promote a more dumbbell shaped distribution, just as we see in Figure 36. The model of lexical evolution simulated here predicts, then, that selection for contrast, cross-category influence, and differential fitness of the available contrast space will interact in such a way that lexicons will tend to exploit fewer less-fit regions of contrast-space more fully, rather than many disparate less-fit regions more sparsely.

2.7 Conclusions

Analogical pressure between categories, whether it arises in the form of cross-category influence in production or perception, in the form of motor program entrenchment (Wedel 2004), or any other mechanism, lies at the heart of the pattern formation effects found in the simulations described in this dissertation. In Chapter 1, simulations were presented in which lexical entries were cyclically updated under analogical pressure in production, producing a steady tendency toward uniform behavior across lexical entries. When context-specific biases were incorporated that introduced conflicting patterns across lexical entries, we saw that analogical pressure resulted in a strong tendency for one pattern to take over and subsume the other, even when the biases on output form were evenly matched. Further, when multiple patterns could conflict, we saw that even though biases operated additively in production, evolved patterns in the lexicon could conform to the outcome of pairwise pattern conflict, in violation of bias additivity. Within Optimality Theory, these two types of conflict resolution correspond to the conflict resolutions resulting from the principles of Constraint Domination, and Strict Constraint Domination, respectively. Within the general model of linguistic competence simulated in Chapter 1, these patterns in conflict resolution do not need to be directly engineered, because they arise spontaneously through more basic interactions over the evolution of the system. This development of higher-order structure through cyclic lower level interactions between individual system elements is known as self-organization.

The simulation results presented in Chapter 2 show that cross-category influence on production interacts with a feedback loop between categorization and production to result in the evolution of a lexicon comprising functionally contrastive categories built from a constrained inventory of elements. When performance biases are introduced to render some elements less-fit, we find that the inventory of sequences evolves to avoid use of those sequences if contrast can be maintained without them. However, if contrast cannot be easily maintained solely through use of fit elements, the lexicon evolves to exploit a subset of the range of less-fit elements. This subset is smaller than expected by chance, reflecting the steady pressure from cross-category influence in production to reduce the inventory to a minimum.

The tendency to reuse features and sequences in concert with the pressure to maintain lexical contrast in this system reproduces the three features of language discussed in the introduction to this chapter, repeated here:

1. Phonological systems do not appear to ever evolve through states in which most or all categorial contrasts are lost.
2. Phonological contrast systems tend to make use of symmetrical inventories, rather than widely scattered contrasts.
3. When a marked element appears in a system of contrast, its frequency is usually lower than that of less marked system elements.

These simulations were based in a model in which perception and production are linked in a feedback loop, where production is more heavily influenced by recent percepts than old percepts. The recency effect of perception in production (as suggested by Goldinger (2000)), was modeled here by the steady replacement of old by new exemplars (Johnson 1997, Pierrehumbert 2001a). In the context of this feedback loop, forms that are sufficiently contrastive to be consistently categorized into one lexical entry have a greater influence on that category's evolution than those forms that are more indeterminate. This indirect selection against less contrastive forms in categorization allows categories to evolve, while maintaining a functionally defined level of distinctiveness, even in the face of a steady tendency to blend forms in production. At the same time, forms do not achieve contrast through completely idiosyncratic use of the available features, because they are constrained by cross-category influence in production.

This indirect selection on the basis of categorization provides a plausible mechanism for perceptual factors to influence the evolution of contrast in lexical entries in interaction with articulatory mechanisms. In the model used here, articulatory biases influence the course of lexical category evolution through steady modification of output forms, which then feedback to the content of categories. At the level of perception on the other hand, this model suggests a mechanism for patterns in confusability to exert an effect at the level of lexical categorization. The model presented here suggests that the conflict between pressure to maintain contrast through selection in categorization, and analogical pressure operating to minimize the number of features and sequences, results in the evolution of a limited inventory of contrasting structures (Lindblom et al. 1984). As a corollary, we expect that contexts in which confusability is lower should tend to evolve to host a greater share of contrasts (e.g., Beckman 1997), while contexts that are universally inhospitable to a given contrast should show a crosslinguistic tendency fail to host that contrast (e.g., as in final obstruent devoicing). However, because contrast is relational, we also expect some patterns to be explicable only by reference to the particular system of contrasts as a whole, as opposed to through solely crosslinguistic considerations (e.g., Lindblom et al. 1984, Lindblom 1992, Flemming 1995, Padgett 2001, 2003).

Finally, because the model described here is non-deterministic, it places no firm boundary on what patterns can possibly become entrenched, but only makes predictions about what kinds of patterns will be more common, and more stable, than others. In this model, the most common patterns that develop across the lexicon will reproduce the effect of some performance bias, given that sequences sensitive to that bias in the lexicon are relatively frequent. However, the positive feedback between analogical pressure and the evolving content of lexical entries can operate over *any* incipient pattern to create and propagate a categorical pattern, allowing incipient patterns that develop simply by chance to also serve as the seed pattern for entrenchment. Crucially, by separating the common sources of seed patterns (in the form of common performance biases), from the mechanism by which patterns can become entrenched in the lexicon, this model can account at once for crosslinguistically common patterns, and for the occasional unusual patterns that develop, using one and the same set of internal mechanisms.

It has long been noted that while phonological patterns come tantalizingly close to reproducing the patterns observed in phonetics, phonologies keep presenting us with their own idiosyncrasies that derail attempts to produce a clean phonetics/phonology mapping (e.g., Bach and Harms 1972, Hellberg 1978, Anderson 1981, Breen and Pensalfini 1999, Mielke 2003, discussed in Anderson 1985 and Blevins 2003). In the model presented here, because the mechanisms by which categorical, productive patterns in outputs arise are functionally independent of the phonetic facts of speech and perception, those facts do not limit what patterns phonologies can reproduce. This model predicts then that a phonological system should in fact be able to reproduce a phonetically unmotivated pattern, such as an alternation between underlying [p] and surface [l] in some context – but it also predicts that such a pattern is unlikely to arise because there is no external bias that can give the system an initial push in this direction (e.g., Blevins (2003).

Rule systems (e.g., SPE, Chomsky and Halle 1968) have been criticized for not being sufficiently restrictive, that is, for being able to describe many patterns that are not, or are only rarely found (e.g., Prince and Smolensky 1993). One might be tempted to call the mechanism for deriving phonological patterns proposed here a Have-Your-Cake-And-Eat-It-Too model, because although it *can* produce many unusual patterns, it is only rarely required by circumstance to do so (e.g., discussion in Mielke 2003). This obtains because analogical pressure acts as an ever-present brake on the entrenchment of distinct patterns. Hence, the only distinctive patterns that are likely to arise are those that either are prompted by persistent biases in performance, or those that emerge in support of contrast between lexical categories. Because of this steady push toward uniformity across categories, the power of the model remains mostly dormant, called upon only when prompted by historical contingency or a statistically rare combinations of events. The result is a model mechanism that is powerful enough to account for the rare, phonetically unmotivated phonological patterns that do exist, but is also constrained by its own internal architecture to conform, *ceteris paribus*, to common phonetic patterns.

Bibliography

Abler, W. L. 1989. On the particulate principle of self-diversifying systems. *J. Social Biol.* Struct.12:1-13.

Albright, A. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language* 78: 684-709.

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

Anderson, S. R. 1981. Why phonology isn't 'natural'. *Linguistic Inquiry* 12: 493-539.

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

Archangeli D. and Pulleyblank D. 1989. Yoruba Vowel Harmony. *Linguistic Inquiry* 20:173-217.

Archangeli D. and Pulleyblank D. 1994. *Grounded Phonology*. Cambridge: MIT Press.

Bach, E. and Harms, R. 1972. How do languages get crazy rules? In R. Stockwell and R. Macauley, eds., *Linguistic Change and Generative Theory*. Bloomington: Indiana University Press.

Batali, J. 2002. The negotiation and acquisition of recursive grammars as a result of comptetition among exemplars. In Briscoe, T. (Ed.) *Linguistic evolution through acquisition*. Cambridge: Cambridge University Press.

Baudoin de Courtenay, J. 1895 (1972). An attempt at a theory of phonetic alternations. In E. Sankiewicz (ed.) *Selected writings of Baudoin de Courtenay*. Bloomington: Indiana University Press.

Baayen, R. H. and Schreuder, R. 1999. War and Peace: Morphemes and Full Forms in a Noninteractive Activation Parallel Dual-Route Mode. *Brain and Language*, 68:213-217.

Beckman, J. 1997. Positional Faithfulness, Positional Neutralisation, and Shona Vowel Harmony. *Phonology* 14: 1-46.

Blevins. J. 2003. Evolutionary Phonology: The emergence of sound patterns. To appear, Cambridge University Press.

Bliese, L. F. 1981. A generative grammar of Afar. *Summer Institute of Linguistics Publications in Linguistics*, 65. Dallas: Summer Institute of Linguistics and the University of Texas at Arlington.

Breen, G. and Pensalfini, R. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30:1-25.

Briscoe, T. (ed). 2002. *Linguistic Evolution through Language Acquisition*. Cambridge: Cambridge University Press.

Broe, M. 1993. Specification Theory: the Treatment of Redundancy in Generative Phonology. PhD thesis, University of Edinburgh.

Browman, C., and Goldstein, L. 1989. Articulatory gestures as phonological units. *Phonology* **6**: 201-251.

Burzio, L. 2001. Sources of Paradigm Uniformity. To appear.

Burzio, L. 2002. Surface-to-Surface Morphology: when your Representations turn into Constraints. In P. Boucher (ed.) *Many Morphologies*, Cascadilla Press. 142-177.

Bybee, J. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: J. Benjamins.

Bybee, J. 1995. Regular morphology and the lexicon, *Language and Cognitive Processes*, 10: 425-455.

Bybee, J. 2000. The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M. and Kemmerer, S. (eds.) *Usage-based models of language*. Stanford: CSLI.

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

Cangelosi A. and Parisi, D. 2002. Computer Simulation: A New Scientific Approach to the Study of Language Evolution. In: *Simulating the Evolution of Language*. Eds: Cangelosi A. and Parisi, D. London: Springer.

Cangelosi A. and Parisi, D. (eds.) 2002. *Simulating the Evolution of Language*. London: Springer.

Chomsky N. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Cohn, A. 1990. Phonetic and Phonological Rules of Nasalization. PhD. Dissertation. UCLA.

Cooper, D. L. 1999. *Linguistic Attractors: The cognitive dynamics of language acquisition and change*. Amsterdam/Philadelphia: John Benjamins Publishing.

Croft, W. 1990. *Typology and Universals*. Cambridge, MA: Cambridge University Press.

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

Cziko, G. 1995. *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution*. Cambridge, MA: MIT Press.

Dawkins, R. 1983. *Universal Darwinism*. In Bendall, D. S. (ed.) Evolution: From Molecules to Men. Cambridge: Cambridge University Press.

Dennet, D. 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Simon and Schuster.

Donegan, P. and Stampe, D. 1979. The Study of Natural Phonology. In D. A. Dinnsen (ed.), *Current Approaches to Phonolgical Theory*. Bloomington: Indiana University Press.

Eddington, D. 2000. Analogy and the dual-route model of morphology. *Lingua* 100:281-298.

Eddington, D. 2002. Dissociation in Italian conjugations: A single-route account. *Brain and Language*, 81.291-302

Ferguson, C. A. 1963. Assumptions about nasals; a sample study in phonological universals. In J. H. Greenberg (ed.), *Universals of Language*. Cambridge, Mass.: 53-60

Flemming, E. 1995. *Auditory Representations in Phonology*. Ph.D. Dissertation. UCLA.

Fodor, J. D. 1981. Does performance shape competence?. *Phil. Trans. Royal Society of London*. vol. B295: 285-295.

Frisch, S. 1996. *Similarity and Frequency in Phonology*. Ph.D. Dissertation, Northwestern University.

Ganong, W. F. 1980. Phonetic Categorization in Auditory Word Perception. *Experimental Psychology: Human Perception and Performance*. 6: 110-125

Goldinger, S. 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, S. 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.

Gordon, M. 2002a. A phonetically-driven account of syllable weight. *Language* 78: 51-80.

Gordon, M. 2002b.Weight-by-positon adjunction and syllable structure, *Lingua* 112: 901-931.

Gould S. J. 1990. *Wonderful Life: The Burgess Shale and the Nature of History*. WW Norton & Co.

Greenberg, J. H. 1966. *Language Universals, with special reference to feature hierarchies*. The Hague: Mouton.

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

Hammond M. 1999. *The phonology of English: a prosodic optimality-theoretic approach*. Oxford: Oxford University Press.

Hayes, B. 1999a. Phonological Restructuring in Yidin^y and its Theoretical Consequences. In Ben Hermans and Marc Oostendorp (eds.) *The Derivational Residue in Phonological Optimality Theory*. Amsterdam: John Benjamins, pp. 175-205

Hayes, B. 1999b. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In Darnell, M., Newmeyer, F. J., Noonan, M., Moravcsik, E. and Wheatley, K. (eds.), *Functionalism and formalism in linguistics,* vol I*, General Papers*. Amsterdam: John Benjamins.

Hayes, B. and Steriade, D. 2003. The phonetic bases of phonological markedness, in Hayes, B., Kirchner, R. and Steriade, D. (eds.) *Phonetically Based Phonology*, Cambridge University Press.

Hellberg, S. 1978. Unnatural phonology. *Journal of Linguistics* 14: 157-177.

Hintzman, D. L. 1986. “Schema Abstraction” in a Multiple-Trace Memory Model. *Psychlogical Review* 93:411-428.

Hurford, J. 1987. *Language and Number*. Oxford: Blackwell.

Hurford. J. 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua*, 77: 187-222.

Hurford, J. 1999. *The evolution of language and languages*. In R. Dumbar, D. Knight, and C. Power (eds.) The evolution of culture. Edinburgh: Edinburgh University Press.

Hurford, J. 2000. Social transmission favours linguistic generalization. In C. Knight., J. Hurford, and M. Studdert-Kennedy (eds.) *The Evolutionary Emergence of Language*, Cambridge: Cambridge University Press. pp. 324-352.

Ito, J. 1988. *Syllable Theory in Prosodic Phonology*. New York: Garland Publishing.

Ito, J. and Mester, A. (2003a). Japanese Morphophonemics: Markedness and word structure. Cambridge, MA: MIT Press.

Ito, J. and Mester, A. (2003b). On the sources of opacity in OT: Coda processes in German. In Fery, C. and van de Vijver, R. (eds.), *The Syllable in Optimality Theory*. Cambridge, UK: Cambridge University Press.

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

Jakobson, R., Fant, G., and Halle, M. 1952. The concept of the distinctive feature. In *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

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

Johnson, K. 1996. Speech perception without speaker normalization. In Johnson, K. and Mullennix J. W. (eds.) Talker Variability in Speech Processing. San Diego: Academic Press.

Jurafsky, D., Bell, A. and Girard, C. 2002. The Role of the Lemma in Form Variation. In Gussenhoven, C. and Warner, N. (Eds.) *Laboratory Phonology* 7. Berlin; New York: Mouton de Gruyter.

Jusczyk, P. and Luce, P. A. 2002. Speech perception and spoken word recognition: Past and present. *Ear and Hearing*, 23:1-40.

Kaufmann, S. A. 1995. *At home in the universe: The search for the laws of self-organization and complexity*. Oxford: Oxford University Press.

Keating, P. 1988. Underspecification in phonetics. *Phonology* 5: 275-292.

Keller, R. 1984 (1994) *On Language Change: The Invisible Hand in Language* (English translation), London: Routledge.

Kenstowicz, M. 1994. *Phonology in Generative Grammar*. Cambridge, MA and Oxford: Blackwell.

Kessler, M. A. and Werner, B. T. 2003. Self-Organization of Sorted Patterned Ground. *Science* 299: 380-383.

Kiparsky, P. 1995. The phonological basis of sound change. In Goldsmith J. (ed.) *The handbook of phonological theory.* Cambridge, MA: Blackwell.

Kirby, S. 1999. *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press.

Kirby, S. 2000. Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., Hurford, J., and Studdert-Kennedy, M. (Eds.) *The Emergence of Language*. Cambridge: Cambridge University Press.

Kirby S. and Hurford 2002. The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model. In: *Simulating the Evolution of Language*. Eds: Cangelosi A. and Parisi, D. London: Springer.

Kirchner, R. 1997. Contrastiveness and Faithfulness. *Phonology* 14: 83-111.

Kochetov, A. 2002. *Production, perception and emergent phonotactic patterns: A case of contrastive palatalization*. New York, London: Routledge.

Kopkalli H. 1993. A phonetic and phonological analysis of final devoicing in Turkish. Doctoral dissertation, University of Michigan.

Krott, A., Baayen, R. H. and Schreuder, R. 2001. Analogy in morphology: modeling the choice of linking morphemes in Dutch. *Linguistics* 39: 51-93

Kruschke, J. K. 1992. ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99:22-44.

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

Labov, W. 1994. *Principles of linguistic change, Vol. 1, Internal Factors*. Oxford and Cambridge, MA: Blackwell.

Ladefoged, Peter. 1984. Out of chaos comes order: Physical, biological, and structural patterns in phonetics. In M.P.R. van den Broecke and A. Cohen, eds., *Proceedings of the ICPhS* **10**: 83-95. Dordrecht: Foris Publications.

Levelt, W. J. M. 1989. *Speaking*. Cambridge MA: MIT Press.

Liberman, M. and Pierrehumbert, J. 1984. Intonational invariance under changes in pitch range and length. In Aronoff, M. and Richard T. Oehrle,

R. T. (eds.) *Language sound structure*. Cambridge, MA: MIT Press: pp. 157-233.

Lindblom, B. 1992. Phonological units as adaptive emergents of lexical development. In: C. A. Ferguson, L. Menn, and C. Stoel-Gammon, eds., *Phonological Development: Models, Research, Implications*, 131-163. Timonium: York Press.

Lindblom, B. 1998. Systemic constraints and adaptive change in the formation of sound structure. In Hurford, J. R. et al., (eds.) *Approaches to the evolution of Language*. Cambridge: Cambridge University Press.

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

Luce, P. A., Goldinger, S. D., Auer, E. T. and Vitevitch, M. S. 2000. Phonetic priming, neighborhood activation and PARSYN. *Perception and Psychophysics*, 62: 615-625.

Luce, P. A. and Pisoni, D. B. 1998. Recognizing Spoken Words: the neighborhood activation model. *Ear and Hearing* 19:1-36.

Maye, J. and Gerken, L. 2000. Learning phoneme categories without minimal pairs. In Howell, S. C., Fish, S. A., & Keith-Lucas, T. (Eds.), *Proceedings* *of the 24th Annual Boston University Conference on Language Development* pp. 522-533. Somerville, MA: Cascadilla Press.

Martinet, A. 1955. *Economie des changements phonétiques*. Berne: Francke.

McClelland, J., L. and Elman, J., L. 1986. The TRACE model of speech perception. *Cognitive Psychology*, 18: 1-86.

McLennan, C. T., Luce, P. A., Charles-Luce, J. 2003. Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 29: 539-553.

Mielke, Jeff. 2003. How to explain natural classes without universal distinctive features. Handout from MOT, University of Toronto, February, 2003.

Moreton, E. and Smolensky, P. 2002. Typological consequences of local constraint conjunction. In Mikkelsen L. and Potts, C. (eds.) *Proceedings of WCCFL 21*. Cambridge, MA: Cascadilla Press.

Nicolas G. and Prigogine I. 1977. *Self-organization in non-equilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley.

Norris, D. 1994. Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52: 189-234.

Ohala, J. 1971. The role of physiological and acoustic models in explaining the direction of sound change. *Project on Linguistic Analysis Reports* (Berkeley) 15: 25-40.

Ohala, J. 1975. Phonetic explanations for nasal sound patterns. In C. A. Ferguson, L. M. Hyman and J. J. Ohala (eds.) *Nasalfest: Papers from a symposium on nasals and nasalization*. Stanford: Language Universals Project. 289 – 316.

Ohala, J. 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, J. 1987. The origin of sound patterns in vocal tract constraints. In P. F. MacNielage (ed.), *The Production of Speech*. New York: Springer.

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

Oliphant M. 2002. Learned systems of arbitrary reference: The foundation of human linguistic uniqueness. In Briscoe, T. (Ed.) *Linguistic evolution through acquisition*. Cambridge: Cambridge University Press.

Oliphant M. and Batali J. 1997. Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11 (1).

Ni Chiosain, M. and Padgett, J. 2001. Markedness, segment realization, and locality in spreading. In Linda Lombardi, (ed.), *Constraints and representations: segmental phonology in Optimality Theory*, Cambridge: Cambridge University Press, pp. 118-156.

Padgett, J. 2001. Contrast dispersion and Russian palatalization. In Hume, E. and Johnson, K. (eds.), *The role of speech perception in phonology*. San Diego: Academic Press.

Padgett, J. 2002. Constraint conjunction versus grounded constraint subhierarchies in Optimality Theory. Ms. University of California, Santa Cruz.

Padgett, J. 2003. The emergence of contrastive palatalization in Russian. In Eric Holt, (ed.), *Optimality Theory and language change*, Kluwer Academic Press.

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

Pierrehumbert, J. 2001b. Why phonological constraints are so coarse-grained. *Language and Cognitive Processes* 16: 691-698.

Pierrehumbert, J. 2002. Word-specific phonetics. In Gussenhoven, C and Warner, N. (Eds.) *Laboratory Phonology* 7. Berlin; New York : Mouton de Gruyter.

Pierrehumbert, J. 2003. Phonetic Diversity, Statistical Learning, and Acquisition of Phonology. *Language and Speech* 46: 115-154.

Pierrehumbert, J. and Beckman, M. 1988. *Japanese Tone Structure*. Cambridge, MA: MIT Press.

Pinker, S. 1991. Rules of Language. *Science* 153:530-535.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., and Slowiaczek, L. M. 1985. Speech perception, word recognition and the structure of the lexicon. *Speech Communication*. 4: 75-95.

Plunkett, K. and Marchman, V. 1991. U-shaped learning and frequency effects in a multi-layered perceptron: implications for child language acquisition. *Cognition*, 38:43-102.

Plunkett, K. and Marchman, V. 1993. From rote learning to system building. *Cognition* 48:21-69.

Prince, A. and Smolensky, P. 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.

Rauch, J. 2002. Seeing around corners. *Atlantic Monthly*, April 2002.

Redford, M.A., Chen, C. C. and Miikkulainen, R. 2001. Constrained Emergence of Universals and Variation in Syllable Systems. *Language and Speech* 44: 27-56.

Saltzman, E. and Munhall, K. G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1: 333-382.

Saussure, F. 1916 (1959). *Course in General Linguistics*, trans. W. Baskin. New York: The Philosophical Library.

Scarborough, R. 2003. Lexical Confusability and Degree of Coarticulation. Paper delivered at the Berkeley Linguistics Society Annual Meeting, 29.

Shadmehr, R. and Bashers-Krug, T. 1997. Functional stages in the formation of human long-term motor memory. *Journal of Neuroscience* 17: 409-419.

Shannon, C. E. 1949. The Mathematical Theory of Communication. In Shannon, C. and Weaver, W. (eds.) *The Mathematical Theory of Communication*. Urbana: The University of Illinois Press.

Shattuck-Hufnagel, S. 1979. Speech errors as evidence for a serial order mechanism in sentence production. In Cooper, W. E. and Walker E. C. T.

(eds.) *Sentence processing: Psycholinguistics studies presented to Merril Garrett*. Hillsdale, NJ: Lawrence Erlbaum.

Skousen, R. 1989. *Analogical modeling of language*. Dordrecht, Boston and London: Kluwer.

Skousen, R. 1992. *Analogy and Structure*. Dordrecht, Boston and London: Kluwer.

Sommer, B. 1970. An Australian language without CV syllables. *IJAL* 36:57-58.

Smith, C. 1997. The devoicing of /z/ in American English: Effects of local and prosodic context. *Journal of Phonetics* 25: 471-500.

Smolensky, P. 1993. Harmony, markedness and phonological activity. Paper presented at Rutgers Optimality Workshop 1. (http//roa.rutgers.edu, ROA# 87-0000).

Stampe, D. 1972. A Dissertation on Natural Phonology. Ph.D dissertation, Ohio State University.

Steels L. 1997. Self-organizing vocabularies. In C. Langton and K. Shimohara (eds.) *Proceedings of the 5th International Workshop on Artificial Life: Synthesis and simulation of living systems* (pp. 179-184). Cambridge, MA: MIT Press.

Steels L. 1998. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103:1-24.

Steels and Kaplan 2002. Bootstrapping grounded word semantics. In Briscoe, T. (Ed.) *Linguistic evolution through acquisition*. Cambridge: Cambridge University Press.

Steriade, D. 1999. Phonetics in Phonology: The case of laryngeal neutralization. In Matthew Gordon (ed.) Papers in Phonology 3 (UCLA Working Papers in Linguistics 2). Los Angeles: Department of Linguistics, University of California.

Steriade, D. 2001. The Phonology of Perceptibility Effects: The P-map and its consequences for constraint organization. Ms., MIT.

Sternberg, S., Wright, C. E., Knoll R. L. and Monsell, S. 1980. Motor programs in rapid speech: Additional evidence. In R. A. Cole (ed.) *Perception and production of fluent speech*. Hillsdale NJ: Lawrence Erlbaum.

Tenpenny, P. L. 1995. Abstractionist versus episodic theories of repetition priming and word identification. *Psychonomic Bulletin and Review* 2: 339-363.

Turner H. 2002. An Introduction to Methods for Simulating the Evolution of Language. In: *Simulating the Evolution of Language*. Eds: Cangelosi A. and Parisi, D. London: Springer.

Tversky, A. 1977. Features of similarity. *Psychological Review* 84: 327-352.

Ussishkin, A. and Wedel, A. 2002. Neighborhood Density and the Root-Affix Distinction. In M. Hirotani, ed., *Proceedings of the Northeastern Linguistics Society* [NELS] 32:539-549.

Ussishkin, A. and Wedel, A. 2003. Gestural Motor Programs and the Nature of Phonotactic Restrictions: Evidence from Loanword Phonology. *Proceedings of the West Coast Conference in Formal Linguistics (WCCFL)* XXII, UC San Diego. March 21, 2003

Wedel, A. 2002**.** Phonological Alternation, Lexical Neighborhood Density and Markedness in Processing. Poster presented at Laboratory Phonology 8, Yale University, New Haven CT.

Wedel, A. 2004. Simulating the evolution of patterns in lexical contrast. Paper given at the Penn Linguistics Colloquium, University of Pennsylvania, Philadelphia.

Worden, R. 2002. Linguistic Structure and the evolution of words. In. Briscoe, T. (ed.) *Linguistic Evolution through Language Acquisition*. Cambridge: Cambridge University Press. pp. 75-110.

Wright. R. 1996. Lexical Competition and Reduction in Speech: A Preliminary Report. In *Research On Spoken Language Processing: Progress Report* No. 21 (1996-1997), Indiana University.

Wright, R. (forthcoming). Factors of lexical competition in vowel articulation. In Local, J. Ogden, R. and Temple R. (eds.), *Papers in Laboratory Phonology* VI. Cambridge University Press.