Self-Organization and Categorical Behavior in Phonology

ANDREW WEDEL
University of California, Santa Cruz

0. Introduction

One of the most salient properties of phonological systems are their very consistency. This paper describes work suggesting that such categorical behavior in phonological systems can be modeled as an emergent property resulting from self-organization (Nicolas and Prigogine 1977) within an iterating, richly-specified lexicon.

1 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 the stipulated mechanism of rule application: if the structural conditions of the relevant rule and/or constraint 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 outcome of all such conflicts.

Where 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. Because the ranking of constraints is fixed, 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. In this paper, I will use the term ‘dominance’ beyond its strict OT usage to refer to a ranking relation between universal 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 follows the results of the component pair-wise pattern conflicts. 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 (Prince and Smolensky 1993). In what follows, I will generalize the OT term ‘strict dominance’ beyond its strict OT usage to refer to the persistence of pattern dominance in the face of conflict with multiple, potentially cooperating patterns.

The aim of the work described here is to show that grammatical relations exhibiting both dominance and strict dominance emerge spontaneously within psycholinguistically plausible models of the lexicon (Tenpenny 1995, Pierrehumbert 2001). The work will be based in a general model of language production and processing that satisfies the following two general conditions:

1) The lexicon is able to simultaneously store detailed information derived from individual events, and multiple categorial abstractions over that information.
2) The mechanism for assembling production targets for a given linguistic element causes such targets to be biased toward the form of other, similar linguistic elements.

The first condition provides a mechanism for low-level phonetic effects to build up and drive category evolution over time, while the second creates a leveling tendency within the lexicon, which will be shown to promote evolution of lexicons exhibiting dominance and strict-dominance patterns in their grammatical relations.

2. Exemplar models of categorization and the lexicon

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). 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 require a form of representation distinct from that encoding specific memories. In the last decade, exemplar theory has 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 (Pierrehumbert 2001).

In exemplar 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 (reviewed in Tenpenny 1995, Hintzman 1986, Goldinger 1996, 2000, Pierrehumbert 2001, 2002)

3. A General Framework

The general framework structuring this work begins with the assumption that every incoming chunk of information can potentially be broken apart and categorized at many levels, such that phoneme level categories coexist with, and are cross-referenced with, categories made up of smaller and larger sequences. In the model, an intention to produce a form, e.g. a word, results in spreading activation of not only previously categorized exemplars of that form itself, but all other exemplars that contain or are contained by it, weighted by strength and number of connections (cf. Pierrehumbert 2001). These activations all may contribute to a production target for that form.

The nature, timing and degree of interactions between categories is left intentionally vague in the present account; for our purposes, the only important feature of the model that derives from these interactions is that a production target be influenced by multiple, cross-categorized elements. This feature has the following important consequence: provided that categorized elements retain their individual values on the dimension that led to their categorization, as in exemplar theory, production of a new category element under any influence of consensus within pre-existing elements will result in ‘blending inheritance’, with consequent reversion to the mean of the category (Abler 1997, Pierrehumbert 2001).

The resulting tendency to assemble outputs that conform to a mean over previously stored forms constitutes, in effect, a form of analogical pressure, operating in this case at the level of phonological production targets. It is this analogical pressure that will be shown in later sections to interact with external biases to produce familiar phonological patterns over time.

4. 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. There are two properties of such systems that will concern us here:

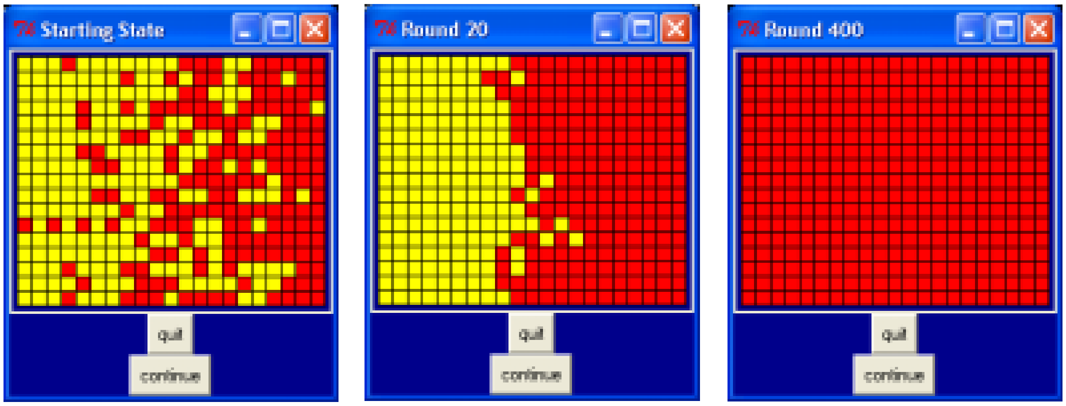
1) Gradient patterns in behavior are unstable; persistent bias towards similarity between elements promotes sharp boundaries in behavior.
2) In the absence of forces that maintain or add difference in a finite system , all system elements will eventually come to exhibit identical behavior.

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 1A-C). At the start of the simulation, each square in a field is randomly assigned a light or dark shade of gray, with the caveat that there is a gradiently higher probability of being assigned dark gray from left to right (Fig. 1A). 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. The result is that the initially gradient light-dark pattern rapidly becomes more categorical, as seen after 20 rounds in Figure 1B.

Note that this segregation occurs despite the fact that there is no explicit directive to do so in the conditions of the simulation: there is only a tendency for a square to become more like its neighbors. This creates a basin of attraction characterized by shade-identity, such that any changes tend to expand consensus neighborhoods 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 seen in Figure 2B.

Because the squares along the boundary have mixed neighbors, they continue to alternate, allowing the boundary to migrate back and forth across the field. When it approaches an edge, the minority shade begins to lose to the majority, and may disappear for good as occurred here on round 400 (Fig. 1C). The initial segregation of shades into two regions, followed by eventual loss of one shade, illustrates a fundamental property of analogical systems: difference is unstable. Here we see that shade difference is initially minimized locally by segregating shades together, but is eventually wiped out globally as well.

Figure 1


A. B. C.

5. Simulation of pattern formation within the framework

The general framework introduced in section 3 is an analogical system, in which production targets for a lexical entry are constructed under the influence of a large set of activated exemplars sharing sequences with that lexical category. The actual productions of lexical entries are consequently biased in the direction of a mean for the lexicon. Re-storage of such productions as exemplars in the lexicon results in a feedback loop, continually updating the lexicon with new exemplars that are more similar to one another.

However, no lexicon ever comes to eventually consist of one word – on the contrary, diachronic change gives the impression of a constantly shifting equilibrium. From the point of view of this framework, such an equilibrium can only be maintained through forces that support or introduce difference within the lexicon. The source of difference that will be relevant in this paper is located in context-sensitive biases in articulation and perception – precisely those gradient, phonetic-level tendencies that so many linguists have proposed form the raw material for grammaticalization (reviewed in Blevins 2003)[1]. The combination of these two general features produces a a system in which idiosyncratic behavior steadily seeps into the lexicon under the influence of low-level performance biases, while reversion to the mean steadily works to reduce that idiosyncrasy. We may predict then that in the ensuing shifting equilibrium, gradient patterns introduced by chance or through consistent biases in performance may occasionally be converted into categorical patterns in the lexicon. The following section introduces a simulation architecture designed to allow testing of this hypothesis within the model.

6. Architecture of the simulation

The architecture presented here is not meant in any way to be taken as a claim about the actual physical functioning of the mental lexicon and production mechanism, but represents rather a serial mechanism for reproducing three key properties that the lexicon and production system are hypothesized to possess, as described above. These properties are

- Multiple nested categorizations of events.
- Influence of multiple categories in assembly of production targets.
- Storage of phonetic detail in memory.

The simulation consists of three parts: a lexicon consisting of abstract lexical entries and corresponding exemplars, an implementation mechanism that uses the information inherent in the lexicon to produce production targets, and a performance filter that introduces biases in actual output form.

In the lexicon, lexical entries are split into two levels, the first an underlying form composed of ordered, abstract phonological categories, and the second a set of more phonetically detailed, remembered exemplars of previous outputs from that lexical entry[2].

---

[1] Pressure to maintain contrast for communicative efficiency provides a distinct set of forces working against leveling pressure in the lexicon. See Wedel (forthcoming) for further discussion.

[2] In the current simulation model, lexical entries are fixed, while exemplars are free to vary under the influence of performance biases and pressure from other exemplars. The disconnect between

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. These summed relations represent a set of weighted implications inherent within the lexicon, each stating essentially, an input substring X should surface as the substring Y. The set of weighted implications is then used to construct a set of possible production targets, where the optimal target enjoys the greatest support over the entire weighted implication set. Once assembled, the production target is sent to Performance.

Performance contains a set biases comprising feature specifications identifying marked sequences, a possible change, and the likelihood of that change occurring. A bias against complex onsets, for example, would be activated whenever a production target contains the feature sequence [onset], [C], and would possibly exchange the feature [coda] for [onset]. The likelihood of this change occurring is set in this simulations here at 10%, meaning that if the production target constructed by the lexicon is [u.bli], there is a 10% chance that it will be articulated as [ub.li] instead. This output is stored in a temporary buffer, and the lexicon begins the process again for another entry. 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. For a more detailed description of the algorithm, see Wedel (forthcoming).

7. Analogical structure results in categorical behavior

The first simulation shown runs over a lexicon consisting of 10 entries of the form VCCV (e.g., /ubli/), seeded with an equivalent number of exemplars with syllable boundaries before and after the first consonant. This simulation includes no performance biases that filter outputs, such that production targets are produced without modification. The symmetry in the syllabification of the seed exemplars means that in the initial 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, resulting in random syllabification choices. However, 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 originating the bias, and from there to the entire lexicon.

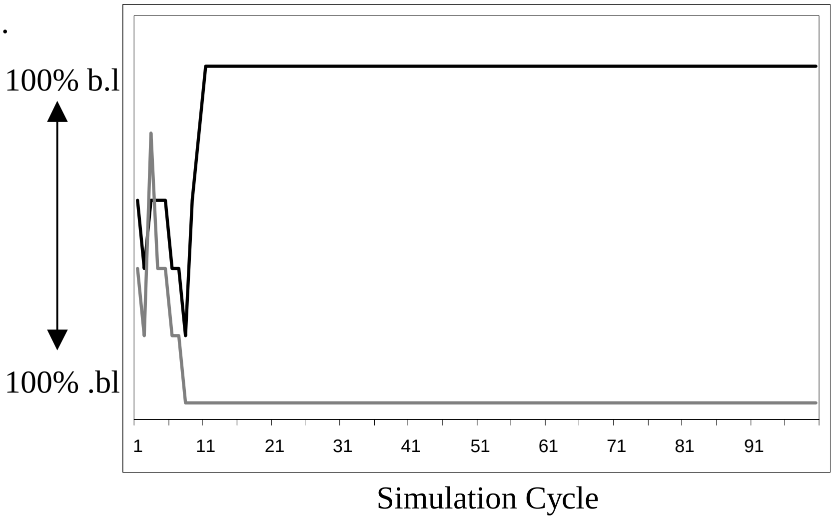
Two different runs with this lexicon are shown in Figure 3. The ordinate

---

abstract lexical entry and exemplar cloud is necessary at this point because the lack of any force maintaining contrast results in a collapse toward uniformity if lexical entries are not fixed. See Wedel (forthcoming) for a version in which lexical entries can freely evolve.

represents the syllabification of production targets from the lexical entry 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 lexical entries 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 lexical entry should rapidly push them to settle on one syllabification or the other. All of the lexical entries share at least some features with one another, however, with the result that many 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 3, where a global syllabification for the entire lexicon is rapidly reached, even if it differed from an early trend within the *ubli* entry. (The behavior of other lexical entries are not shown here, as they all reach a consensus syllabification at approximately the same time.)

Figure 3.



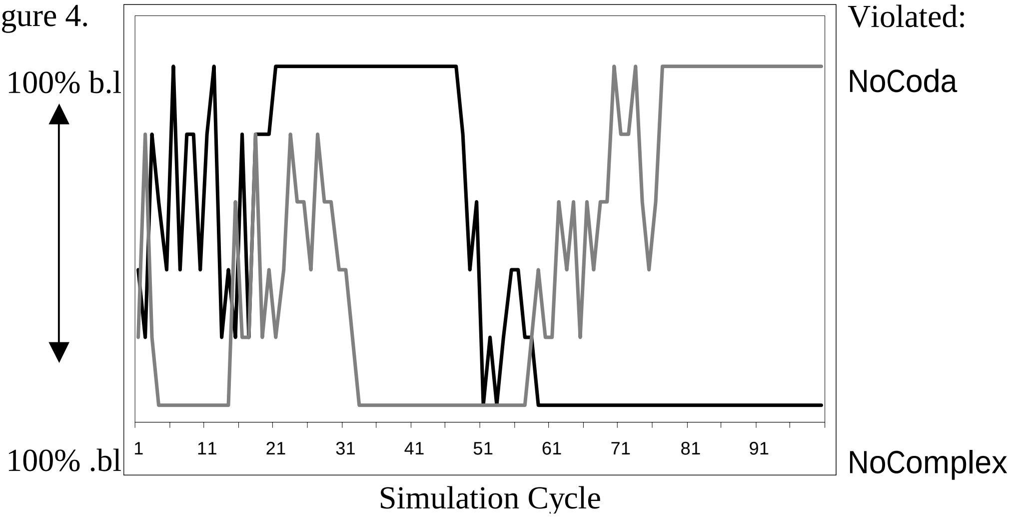
Note that the two runs shown converge on opposite syllabifications. Although there is no way to predict ahead of time which syllabification will sweep the lexicon, because the system is analogically structured, we *can* predict that sooner or later one syllabification will indeed win out. This is precisely analogous to the cellular automata simulation shown in Figure 1, 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.

8. Addition of external noise results in oscillation between extremes

The following simulation is run with the same lexicon shown in Figure 3, 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 4, 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.

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. 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.

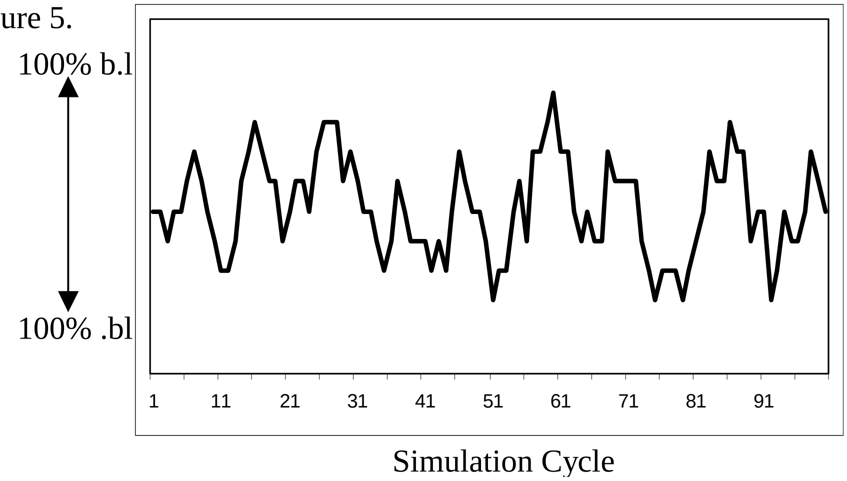
Figure 4.



9. Categorical behavior is dependent on analogical structure

To show that categorical behavior is in fact crucially dependent on analogical pressure in this system, the next simulation is 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 5, show precisely what we expect in the presence of evenly matched, contradictory biases: syllabification behavior that simply oscillates around the mean.

Figure 5.



10. The development of patterns conforming to strict 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, i.e. that the choice of optimal outputs proceeds through satisfaction of constraints in strictly ranked order. The principle of strict domination, in particular, specifies that ranking is absolute: no degree of potential violation of lower ranked constraints can ever compel violation of a higher ranked constraint. These limitations allow OT to predict that certain patterns cannot exist.

However, while strict domination allows OT to accurately describe many phonological systems, it sits uneasily with the notion, held by many, that constraints in OT are directly or indirectly related to physical factors external to the grammar. This unease arises because it is difficult to see how physiological constraints on articulation or perception would not interact additively in some overall performance cost.

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 (Gordon 2002). 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 principle of OT.

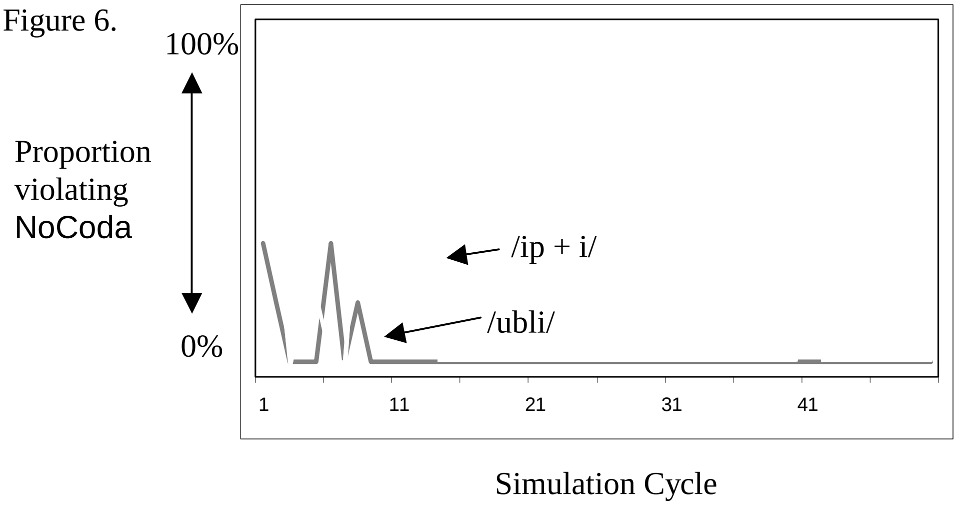
11. Setting up multiple constraint conflict in the simulation

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 has 8 entries in each class with 5 exemplars each.

The performance filter is outfitted with three biases, NoCoda and NoComplex, and a third, abbreviated AlignStem, which operates to move syllable boundaries to coincide with stem boundaries. For the ‘ubli’ class, the two syllabifications trigger NoCoda or NoComplex, while for the ‘ip + i’ class, the two syllabifications trigger NoCoda or AlignStem, respectively. Crucially, 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. This situation gives us a chance to look 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 biases both contribute to the total ‘badness’ of that syllabification. Hence, these biases should jointly contribute to the pressure to grammaticalize the alternative, NoCoda violating syllabification.

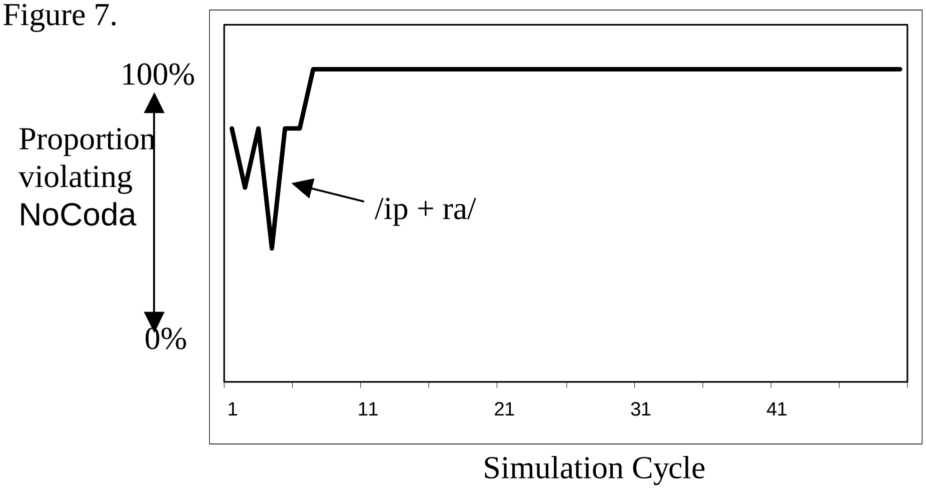
To give these biases an opportunity to do so, the simulation will operate with the bias strengths set such that NoCoda (15%) is a stronger bias than either NoComplex (10%) or AlignStem (10%) alone, but that the latter two together outweigh NoCoda. To show that these relative bias strengths do in fact result in the expected biases in output form, I show typical results of a simulation run over a lexicon containing just the /ubli/ and /ip + i/ word classes in figure 6, and a simulation run over a lexicon containing the /ip +ra/ class in isolation in figure 7.

Figure 6.



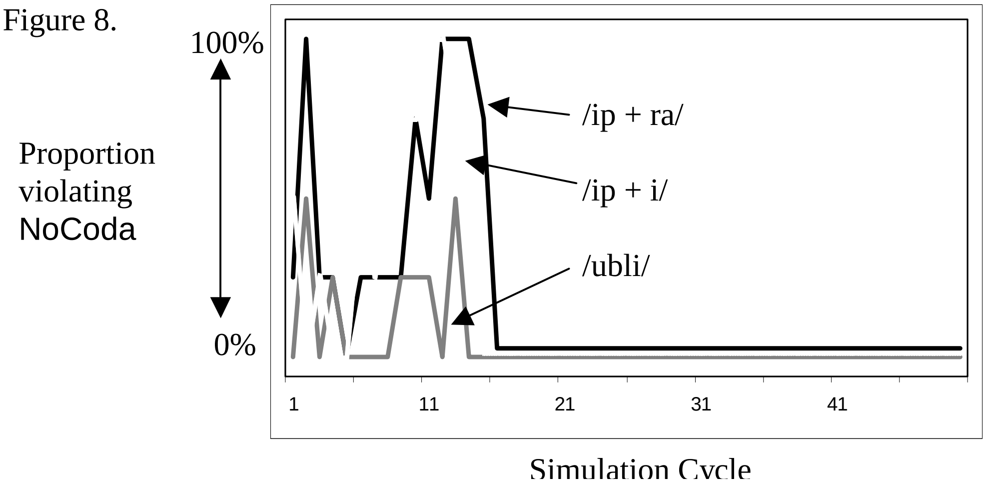
So far then, we see that when the lexicon contains only lexical entries of the /ubli/ and /ip + i/ classes (Fig. 6), the NoCoda pattern dominates the possible NoComplex and AlignStem patterns. On the other hand, when the lexicon contains only lexical entries of the /ip + ra/ class (Fig. 7), the joint NoComplex/AlignStem pattern dominates the possible NoCoda pattern. A representative simulation with a lexicon containing all three classes *together* is shown in figure 8. As before, the /ubli/ and /ip +i/ classes evolve to categorically

Figure 7.



assemble production targets without codas. However, we find that after some hesitation, the /ip + ra/ class does likewise, contradicting its behavior when evolving alone.

Figure 8.



The crucial difference between the simulation in figures 7 and 8 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 that class, but also influenced indirectly by performance's bias for forms *without* codas in the other two classes.

12. Conclusions

The simulations results presented above support the prediction that domination of one possible phonological pattern over another will be a structurally defined 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, reversion to the mean sets up positive feedback loops over production and perception with the result that categorical dominance, both relative to pairs of biases, and to larger interacting bias sets, is a spontaneously reached, stable evolutionary state. Neither domination nor strict constraint domination need to be stipulated in the model, as they are straightforward consequences of lexicon structure and the mechanism by which production targets are assembled.

References

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

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

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

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.

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

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

Johnson, K. 1997. Speech perception without speaker normalization. In Johnson, K. and Mullennix J. W. (eds.) Talker Variability in Speech Processing. San Diego: Academic Press.

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

Pierrehumbert, J. 2001. 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. 2002. Word-specific phonetics. In Gussenhoven, C and Warner, N. (Eds.) *Laboratory Phonology* 7. Berlin; New York : Mouton de Gruyter.

Prince, A. and Smolensky, P. 1993. *Optimality Theory: Constraint interaction in generative grammar*. ms. University of Colorado at Boulder CO and Rutgers University, New Brunswich, NJ.

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

Wedel, A. (forthcoming). Self-Organization and the development of higher-order phonological patterns. Ph.D Thesis, University of California, Santa Cruz, CA.

Department of Linguistics
University of Arizona
1100 E. University Blvd.
Tucson AZ 85721