Implications of an Exemplar-Theoretic Model of Phoneme Genesis:

A Velar Palatalization Case Study*

Running Head: Phoneme Genesis

Rebecca L. Morley
The Ohio State University
Columbus, OH

* This work was supported by a Targeted Investment in Excellence from the OSU College of Humanities. I would like to thank Andrea Sims for attempting to explain Slavic palatalization to me. Thanks to the Phonatics group at Northwestern for their feedback on an earlier version of this paper. Thanks also to William Schuler, and especially Fred Mailhot, for their helpful comments. This paper has been greatly improved by the thoughtful suggestions of Jeff Mielke, Juliette Blevins, and James Scobbie.

Address all correspondence to Rebecca L. Morley, The Ohio State University, Dept. of Linguistics, 200 Oxley Hall, 1712 Neil Ave., Columbus, OH 43210 Email: morley@ling.osu.edu

ABSTRACT

Diachronic velar palatalization is taken as the case study for modeling the emergence of a new phoneme category. The spread of a palatalized variant through the lexicon is treated as a stochastic classification task for the listener/learner. The model combines two measures of similarity to determine classification within an exemplar-theoretic framework (Nosofsky, 1988): acoustic distance (e.g., Pierrehumbert, 2003), and phonotactic expectation (e.g., Dupoux, Kakehi, Hirose, Pallier & Mehler, 1999). There are three model outcomes: contrast, allophony, or contextual neutralization between the plain and palatalized velars. It is shown, through a series of simulations, that these can be predicted from the distribution of sounds within the pre-change lexicons, namely, the ratio of the /k-vowel/ sequences containing naturally palatalizing vowels (i, ɪ, e), to those containing non-palatalizers. "Unnatural" phonotactic associations can arise in individual lexicons, but are sharply limited due to the large size of the lexicon and the local nature of the phoneme changes. "Anti-Natural" distributions which categorically violate the proposed implicational relationship between palatalization and frontness/height (Bhat, 1978; Neeld, 1973) are absent. This work provides an explicit and restrictive model of phoneme change. The results also serve as an existence proof for a non-UG based mechanism of avoiding over-generation.

Key words: Exemplars, Sound Change, Simulation, Universals, Evolutionary Phonology

INTRODUCTION

This paper describes an exemplar-based computational model of language change. The model is developed in order to test the predictions of a general class of emergentist theories of linguistics. We ask if the range of synchronic grammars produced by running the model repeatedly under different conditions is an accurate reflection of the known typology. One objection to the emergentist approach stems from the belief that such models would over-generate unnatural and anti-natural grammars – unattested patterns that do not conform with posited language universals (de Lacy, 2006; Kingston & de Lacy, 2006; Kiparsky, 2006, 2008).

The model chosen is one that is judged to be among the simplest instantiations that still includes the elements necessary to test the theoretical questions of interest. A dissimilation-based sound change supplies the necessary non-phonetically motivated segment sequences. The element of chance is provided by a sufficiently large lexicon to which sound changes are stochastically applied. And a top-down pressure from a phonotactic grammar allows fortuitous segment associations to persist. The model adopts a basic exemplar framework where individual word and sound tokens are stored in memory and directly affect perception (categorization) and production (e.g., Bybee, 2002; Goldinger, 1996; Luce, 1986; Nosofsky, 1988; Pierrehumbert, 2001). The Phonotactic grammar, in turn, is assumed to derive from computations over the abstract categories to which the tokens belong.

A series of simulations using this model elaborate the historic conditions under which certain types of synchronic segmental relationships will arise: contrastive, allophonic, and neutralizing. Additionally, it is found that this simple – but, in certain ways, sufficient – model produces a relatively narrow typology; it does not significantly over-generate. This provides a concrete example – an existence proof – of a means of restricting synchronic grammars which does not rely on innate substantive restrictions.

Additionally, this work brings together usually disparate strands of research involving historic change and diffusion, exemplar theory, synchronic phonological representations, and the methodology of computational modeling. In so doing, the assumptions necessary to create a consistent and coherent theory of the transition from phonetic variation to phonological alternation are highlighted. Rather than provide definitive answers to long-standing questions within theoretical linguistics – both synchronic and diachronic— this work provides clear predictions which can be traced to particular features of the chosen model.

The paper is organized as follows. In the next section the case study for the investigation of these questions – velar palatalization – is described in detail. An introduction to the framework and terminology of sound change which are adopted precede an in-depth description of the theoretical and computational components of the model. The results of the simulations are discussed in the following section, with an examination of which facets of the model are responsible for the distribution of outcomes, and which initial lexical conditions lead to which particular synchronic grammar. The remainder of the paper looks in detail at the possible scenarios under which synchronic systems which violate supposed palatalization universals *could* arise from an emergentist framework. It is possible to construct such a trajectory within the parameters of the model. However, given the large number of necessary conditions, it is estimated that the likelihood of such trajectories are low. Finally, the paper is concluded with a summary and discussion of the findings, and the implications of these for both synchronic and diachronic theory.

THE CASE STUDY

This paper will investigate specifically a case of phoneme split where a novel palatalized velar phoneme emerges over time, evolving from a plain velar precursor. Velar palatalization is chosen largely because it is well-documented both diachronically and synchronically – taking as examples of the same phenomenon cases in which variants surface as palatal affricates, palatal-alveolar affricates, dental or alveolar affricates, etc. Synchronically, there are a number of attested patterns in which plain velars (e.g., /k/) and palatalized variants alternate in their environments: the latter occurring only in the context of certain high front vowels, the former occurring elsewhere. In Lamba, a Niger-Congo language spoken in Zambia, for example, /i/-initial suffixes trigger palatalization on stem-final /k/'s, as in (1) (Odden, 2005). Similarly, in Serbo-Croatian, an alternation between velars and palatal-alveolars can be seen in the comparison between the 1st person singular present tense, and the three genders of the past tense in (2) (Kenstowicz, 1994).

(1) Lamba (Odden, 2005)

/ʃik+ika/→	*[ʃitʃ-ika]* bury-NEUT 'bury'
/kak+ila/→	*[katʃ-ila]* tie-APPLIED 'tie'

(2) Serbo-Croatian (Kenstowicz, 1994)

1st Sing. Present	Masculine Past	Feminine Past	Neuter Past	Gloss
petʃém	pékao	peklá	pekló	bake
ʒeʒém	ʒégao	ʒeglá	ʒegló	burn

Kochetov (2002) lists Russian, Czech, Polish, Slovak, Ukrainian, Bulgarian, Lithuanian, Scots Gaelic, Nenets, and Japanese, among others, as exhibiting some type of plain/palatalized alternation. Phoneme changes which convert plain velars to a palatalized variant are also well known, and are common within the Slavic family; Guion (1998) additionally cites Indo-Iranian, Cowlitz Salish, Bantu, Old English, Mam, and Old Chinese. It has been argued that a fronted tongue position in the presence of a front vowel leads to acoustically similar (and confusable) productions for sequences of /ki/ and /tʃi/ (Guion, 1998; Keating & Lahiri, 1993; Winitz, Scheib & Reeds, 1997), providing a possible misperception source for the phoneme change.

Finally, theoretical work on the nature of linguistic universals has led to the proposal for an inviolable implicational hierarchy (based on work by Bhat, 1978 and Neeld, 1973). This formulation takes the observed typology as the result of a scale of harmony, such that k^ji sequences are preferred to k^jɪ sequences, are preferred to k^je sequences and so on. The entire scale for the 7-vowel inventory that will be used for this paper's model can be written in the following way.

(3) Universal Harmony Scale: k^ji > k^jɪ > k^je > k^jo > k^jʊ > k^ju > k^ja

Within Optimality Theory (Prince and Smolensky, 1993/2004), this scale embodies the implication that if a language allows k^je sequences, it must also allow k^ji sequences, and the same is true for any two elements in which the first is to the right of the second on the scale.[1] Such implicational scales are often taken to apply to contrasts as well, in the reverse direction, such that if a language tolerates an auditorily close contrast between ki and k^ji (i.e., a hard to discriminate distinction), it will also tolerate any less close contrasts, say between k^je and ke. The output of the model will be tested with respect to these universal claims.

PHONEME CHANGE

[1] This scale is derived from the standard IPA vowel-space positions, and is assumed to be a decreasing function of both frontness and height. Language-specific instantiations of particular vowels may result in slight variations, such as a more front location for /u/ which will enhance its palatalizing degree (see Guion, 1998). /a/ is assumed to surface as a low central vowel. In practice, the scale is taken to apply only to the first three vowels since /k/ might actually be expected to be pronounced as more back, or velarized, before back vowels. In the model, a phonetic palatalization effect only applies to i, ɪ, and e.

Before proceeding any further it is necessary to lay out the theoretical perspectives and terminology adopted in this paper. For a Neogrammarian the phrase 'sound change' refers to a gradual change in the phonetic instantiation of a particular phoneme. Although the change is gradual, it is regular in the sense that all tokens of the given class are shifted to the same degree. Theories of lexical diffusion, on the other hand, ascribe gradualism not to the degree of phonetic shift, but to the participating tokens. An abrupt sound change will spread through a lexicon gradually. Once all words containing the segment conform to the new phonetic instantiation the change is complete.

A variationist perspective assumes that all potentially regular sound changes exist, in embryonic form, within synchronic systems. Productions vary within and across speakers, and across speech communities. Particular variants are chosen – or exaggerated – to mark community membership (see Guy, 1980, 2008). Changes in the relative distribution of different idiolects within a language community can be modeled as a competition between existing grammars (see Niyogi & Berwick, 1997). However, phonological rules and phoneme classes are invariant. This is true of all three frameworks. Although each can be said to capture some aspects of diachronic change, none is designed to deal with changes to representations.

In order to describe change in the phonology of a language there must be some mechanism for transitioning from phonetic to phonemic change. In their 'big bang' theory of sound change Janda and Joseph (2003) (see also Janda, 2008) suggest that a phonetically based sound change is phonologized once the socially marked variants become phonetically dissimilar enough, although the exact mechanism for this is not specified. An earlier view, which they argue against, posits that new phonemes are created when existing allophones lose their conditioning context (see e.g., Hajek, 1997; Kiparsky, 1995). The problem with this view as they, and many others, point out, is that loss of context should cause loss of the allophonic features conditioned by that context as well. To be consistent with the over-all theory, the only way for those features to be retained is if they have already been lexicalized. That is, the loss of the trigger can occur *because* phonologization has already taken place. This account, unfortunately, lacks an explanation for the original phonologization event.

This work will be most concerned with changes in abstract representations, changes that are discernible through distributional evidence and that reflect shifts in mental representations. We will therefore avoid the ambiguous 'sound change' in favor of the term 'phoneme change'. Phoneme change will refer specifically to a change in the underlying phoneme class of a given phonetic token.[2] This change will be assumed to occur separately from phonetic change, although the two will naturally be correlated.

From the Neogrammarians, the model will borrow phonetic gradualism, and from the diffusionists, lexical gradualism. Along with the variationists, and the usage-based approached (Aski, 2001; Bybee, 2001, 2002, 2008), the synchronic grammars of this model will contain numerous phonetic tokens conditioned by linguistic as well as social context.[3] It is not the goal of this work to directly resolve the debate over the actuation problem; rather, by assuming the initiation of a new phoneme category, we will be able to focus on the internal evolution of that category within a synchronic system. The way the language as a whole can be said to change through the changes to the structure of this category will be instructive for thinking about the factors that can lead to – or suppress – the genesis of novel phonemes.

THE MODEL

The work in this paper straddles the diachronic and synchronic domains, explicitly linking the two ends of the phonetics-phonology continuum. At one end there is a large body of work on

[2] Assuming an explicit phoneme level of categorization makes the discussion more straightforward, but is not necessarily critical to the argument. A model in which phoneme-like behavior emerges from structures at higher levels may be adopted as well (see, e.g., Wedel, 2004).

[3] Sociophonetic variation is not modeled directly. However, individual tokens are subjected to random variation in phonetic space. This can be used as a stand in for the known degree of variation due to factors such as social marking. Although the sociological aspects of language change are not the focus of the current work, the model is compatible with a framework in which social variation is the impetus for language change. The individual speaker, as modeled in this paper, can be placed in a larger network of agents where communicative pressures will act. The point is not to replace social factors with other mechanisms, but to focus on internal factors, which we assume can be productively studied in isolation.

how individual words and segments are affected by shifts in pronunciation and how those shifts might incrementally be adopted within a given lexicon or language community (e.g. Bybee, 2002; Kinkade, 1973; Labov, 1972, 1981;Yeager, Labov & Steiner, 1972, and many others). At the other end, there is a considerable apparatus for determining how grammars can be learned from input data, and under what conditions a given hypothesis will be preferred to a competitor (e.g., Albright & Hayes, 2003; Boersma & Hayes 2001; Gibson & Wexler, 1994; Jarosz, 2006; Pearl, 1997; Pearl & Lidz, 2009; Tenenbaum & Griffiths, 2001; Tesar & Smolensky, 1998). What is missing is the link between the two ends: explicit modeling of the effect of sound and phoneme changes on the data available to the learner, and the effect learning over such data will have on synchronic grammars.

The first contribution of the modeling work of this paper is in drawing attention to unspecified elements within current theories. Implementation of such a comprehensive ("soup to nuts") model forces explicit decisions about the form of a similarity and distance metric between phonetic exemplars. It also forces assumptions about the range of natural (that is, phonetically motivated) phoneme changes. These are parameters that must be set in order to produce concrete theoretical predictions, and they are gaps in our knowledge that must ultimately be filled by further research.

Figure 1 illustrates the basic model elements. Changes in phoneme classification (due to the Source) are assumed to apply to individual sounds as perceived by a listener/learner. These sounds, in turn, comprise words; thus changes in a particular sound token change the phonological properties of the word in which they appear. These words, in turn, comprise a speaker/listener's lexicon. This entire set of words is taken as the data from which the learner induces their grammar. The grammar, for the purposes of this paper, is a series of conditional probabilities which encode surface restrictions on preferred and dispreferred phoneme sequences – a phonotactic grammar.

Figure 1 approx here

The grammar affects future phoneme changes in the following way: by boosting the probability of changes that would result in preferred sequences, and decreasing the probability of changes that would result in dispreferred sequences. The resulting lexicon is then used to update the phonotactic grammar. This process can continue indefinitely. The interaction of these two mechanisms is shown by the loop in Fig 1. There are no 'completed' sound changes in this model, but halting the simulations at a given point in time provides a sample output Grammar.

There are, by hypothesis, two distinct mental categories to which tokens can belong: the plain velar, and the palatalized velar. These may overlap in phonetic space completely, partially, or not at all, at any given iteration of the model. This property is taken to be distinct from the segmental distribution which is used to characterize the relationship of the two phonemes as allophonic, neutralizing, or contrastive. The latter type of relationship will be used to characterize the predictions of the model.

The full specification of the model components and the accompanying working assumptions will be given in the following sections. Their salient properties, however, can be summarized here as follows. The model is designed to be as simple as possible while retaining cognitive plausibility in critical areas. One of the basic questions under test in this paper is whether, in order to reproduce the known typology, a given hypothetical learner will need to have additional guidelines *a priori* about which grammars should *not* be considered. To test this, the current model is missing any such guidelines – the Learner is substantively naïve.

The Source is designed to contain a representative set of phoneme changes, critically, ones that both reinforce natural patterns (assimilatory), and ones that erode them (dissimilatory). The Lexicon is large enough, and varied enough, to generate spurious statistical associations that the learning mechanism can discover. This combination of factors provides routes for unnatural, and theoretically undesirable, grammars to arise.

The Source

The particular phoneme changes described in this paper are all ones that have been described as occurring within the family of Slavic languages, although not necessarily all within the same language, or in the exact sequence described here. Explanations, when offered, are varied, often making reference to functional or cognitive factors (Andersen, 1978; Fairbanks, 1965; Hamilton, 1980; Matthews, 1960; Shevelov, 1979; A. Sims, personal communication, August 20, 2009; Townsend & Janda, 1996; Zilynskyj, 1979). The work in this section re-casts the selected set of phoneme changes within a single framework for the purpose of specifying the characteristics of the Source. This is done, in particular, with an eye towards isolating the general factors that shape the evolution of emerging phoneme categories.

The methodology is as follows. We stipulate a relatively simple and general model of internal phoneme split. This model is then incrementally expanded to the point where it can specify a complete trajectory from phonetic correlation to phonological representation. The computational approach requires that there be no ambiguity about how forms are processed at any stage of the model. As a result, we will be required to commit to a number of hypotheses about aspects of phoneme change and language learning about which little is known. It should be kept in mind that these working hypotheses are provisional, placeholders for facts that will be supplied by future research.

Table 1 contains the full set of phoneme changes that will be implemented in the simulations to follow. Stage 0 begins with the phonetic palatalization of velars before high, front vowels. Although it has been shown that the degree to which phonetic processes are actualized is a language-specific property (e.g., Beddor, Harnsberger & Lindemann, 2002; Keating, 1985; Solé, 1992), universal properties of the human articulatory and perceptual systems are generally taken as the basis for these processes. Thus, phonetic palatalization (of some particular degree) can be attributed to a combination of articulatory overlap (e.g., Cutler, 1987; Lindblom, 1983), and auditory ambiguity in the signal, both of which act to merge adjacent segment features (Guion, 1998).

Within the current model, jer deletion (Stage 1) is taken as the event that initiates the formation of the novel /k^j/ category. This category is seeded by the set of words that lose a final [ɪ] vowel in the presence of a preceding /k/. The contrast is established, by hypothesis, by the lack of predictability that results from the jer loss – producing minimal pairs which would otherwise become homophones (e.g., /dak^j/ from /dakɪ/, and /dak/ from /dakʊ/, rather than /dak/ from both).

In the remainder of the paper, the subsequent internal evolution of this newly minted category – and its parent, the /k/ category – is charted within the grammar of an individual speaker. Whether such a novel contrast will spread throughout an entire language community will depend on a number of additional factors, many of them social. That process will not be modeled here. The primary focus of this work is how an internal change will interact with an existing lexicon and an existing grammar.

Although a communicative interaction is not explicitly modeled, there is an implicit speaker from whom the learner receives input. Input tokens are variably categorized based on perceptual and cognitive factors. This is functionally quite similar to a scenario in which a listener is exposed to forms from multiple speakers with distinct idiolects. Additionally, as in multi-generational models, outputs at one stage become the inputs for the next, such that successive changes can accumulate (e.g., Daland, Sims & Pierrehumbert, 2007; Kirby, 2001). Where this model differs from other diachronic dynamical modeling work is in the system being modeled. Typically, the focus is on the evolving distribution of pre-existing idiolects in a particular population (e.g., de Boer, 2000; Hutchins & Hazlehurst, 1995; Niyogi & Berwick, 1997; Regier, 1995; Steels, 2000, 2005). The relevant dynamical system for the work in this paper, however, is the lexicon-grammar pair, and the evolving distribution of the phonemes within the words of that lexicon, and from which the grammar is computed.

Table 1 approx here

At the beginning of stage 2 a large /k/ category and a smaller /k^j/ category share phonetic space. /k^j/ tokens are in especially close phonetic proximity to tokens of /k/ that occur before high, front vowels. This proximity, or similarity, is the source of potential miscategorization within a listener-oriented model of phoneme change (see Blevins, 2004; Fujimura, Macchi & Streeter, 1976; Hura, Lindblom & Diehl, 1992; Ohala, 1981, 1990, 1993; Repp, 1977). At Stage 2, /k/'s that are phonetically palatalized before front vowels may be mistaken for /k^j/'s – assimilating features of the adjacent vowel. The reverse type of dissimilatory change is also possible, in which newly minted /k^j/'s revert to /k/'s when listeners over-compensate for the expected phonetic influence of the neighboring vowel.

It is worth pausing here to draw attention to the logical requirement that formation of the /k^j/ category precede the miscategorization of individual /k/ sound tokens. Final jer deletion is responsible for the creation of that category in the current model, but it may be thought of as a stand-in for a special kind of event that acts to create novel category labels.[4] In the case of phoneme split (as opposed to merger, or neutralization) the existence of a competitor category cannot be taken for granted, and must be accounted for in some way. It is also important to be clear about the role that mental categories play in linguistic theories, especially ones within an

[4] Treating the initial jer deletion as a regular and complete phoneme change is, in fact, antithetical to this type of modeling. Such a change should be treated in exactly the same way as the palatalization changes. The reasons for not doing so are entirely pragmatic. In initiating the model with a point event we avoid a sort of infinite regress of partial, probabilistic and context-dependent changes. Ignoring any complexities related to vowel loss, we are able to focus entirely on the palatalization question. It is always the intention, however, to build the model up gradually to the point where simplifications such as this one can be dispensed with.

exemplar framework. There is sometimes a mistaken assumption that exemplar theory does away with categories. On the contrary, categories are an inherent element of any exemplar theory. Exemplars are not defined unless they are the exemplars of something. Even models lacking a phonemic level of representation require categories to exist at some other level, such as the word, or the utterance. Furthermore, any exemplar model that deals with a competition in categorization, or a shift in pronunciation, does so via assignment of category labels to distinct sets of exemplars (although category membership may be probabilistic).

The stochastic application of the paired set of phoneme changes at Stage 2 allows us to fold in factors – such as word frequency, or social context – which are known to affect phoneme change without modeling them explicitly (see, e.g., Bybee, 2001, 2008; Labov, 1981; Phillips, 1984; Pierrehumbert, 2002; Weinreich, Labov & Herzog, 1968). Phonetic context and phoneme change type will be explicitly modeled. The likelihood with which a plain /k/ converts to a palatalized variant will depend on the palatalization strength of the following vowel: /i/ is a stronger palatalizer than /ɪ/, and /ɪ/ is stronger than /e/ – see the harmony scale in (3). Following Guion (1998)'s finding of a perceptual asymmetry between the two directions of change, the likelihood of $k > k^j$ is taken to be higher than $k^j > k$ (see also Chang, Plauche & Ohala (2001) for similar results).

It is further assumed that once the triggering event (final jer deletion) has occurred, re-categorization will be a constant facet of speech processing. This is not a necessary element of the misperception change model, but it allows historic phoneme change to be instantiated incrementally through iterations over synchronic systems. This is useful for simulation purposes as it allows fine-scale sampling of the interaction between individual word tokens and the system of words and sounds to which they belong.

The final phoneme change, at Stage 3, is included as representative of a type of dissimilatory change. The existence of two acoustically similar phoneme categories that must be differentiated is posited to exert a pressure on the articulation of words containing both phonemes (a hyper-articulation change; see Ohala, 1981). Again, this change is critically dependent on contrast – something that is impossible without the existence of at least two categories. In the example in Table 1, in words like /k^jek/, the intermediate vowel develops a diphthongal structure; the initial part of this diphthong is then re-analyzed as a perceptual artifact of the underlying palatal variant: /k^j/, resulting in the final /k^jok/ form in Stage 3. The inclusion of this type of phoneme change also introduces environments for the palatalized velar which are not high, front vowels, and thus, not 'natural'. This distinction will become important in the discussion of phonological universals below. The properties of the Source are summarized in (4).

(4) The Source

- Phoneme changes are based in universal phonetics
- Changes result from misperception and/or misanalysis on the part of the listener
- Changes apply stochastically by segment token
- Individual phoneme changes are either assimilatory or dissimilatory by local phonetic context
- Changes are always active (in both directions)

The foregoing discussion serves to specify aspects of a working theory of internal phoneme change. This explicit characterization allows us to make concrete predictions about language change. Application of the set of changes in Table 1 – to varying sets of words, and for a given learner – produces a set of grammars. If all parameters have been exhaustively varied, this set describes the space of possible grammars discovered by a naïve learner exposed to the products of exemplar-based phoneme change. Those predictions can then be tested against the existing typology to determine if the model is incomplete or incorrect in some way.

The Lexicon

The phoneme changes described in the previous section act on the sounds of particular words – words that belong to the lexicon of a speaker of a particular language at a particular point in time. In Table 1 a handful of example words were used to illustrate the relevant phoneme changes, beginning with the pre-change pair /dakɪ/ and /dakʊ/. The outcome of this series of changes cannot be directly determined from this table however; this is because changes are stochastic, act in opposing directions, and are affected by prior language experience. Since the phonotactic grammar modeled here is induced from the words of the lexicon, the outcome crucially depends on the exact nature of the lexicon. A new lexicon, therefore, was generated for each set of model simulations. These lexicons are meant to be more or less representative of real-world languages. Words varied with respect to syllable structure and length, with a large enough inventory of vowels to represent the scale in (3), and a small enough set of consonants such that spurious inter-segment correlations could arise. The lexicon makeup is described by the specifications in (5).

(5) The Lexicon

- The consonant inventory consists of {p,t,k,b,d,g,s,z,m,n}
- The vowel inventory consists of {i,ɪ,e,o,ʊ,u,a}
- The syllable type inventory consists of {CV,CVC}
- Each lexicon contains 6000 words of three types: {1-syll,2-syll,3-syll}

Initially, only words containing the sequence kɪ# (where # indicates the end of the word) are possible contexts for a palatalizing phoneme change (Stage 1 of Table 1). Once this change has occurred, word-internal ki, ke, and kɪ sequences are susceptible to change as well, and Stage 2 and 3 phoneme changes are possible. Table 2 provides an example of a subset of words taken from a particular lexicon at five sequential stages in the model (from left to right). Bolded font indicates that a change in the underlying form has occurred from one stage to the next. Gray cells indicate words which had the potential to change (i.e., met the conditions of the phoneme change in that column), but failed to do so. A gray cell in column 2 (Stage 1) indicates that, although jer deletion occurred, the pre-vocalic velar failed to retain palatalization (a fixed 75% rate of palatalization was used).

Table 2 approx here*

The Grammar

From the words of the lexicon the model learner directly computes a set of conditional probabilities over pairs of segments. This constitutes their phonotactic grammar (see, e.g., Hayes & Wilson, 2008). At each application of a possible phoneme change the learner's grammar affects the likelihood of that change. In the reverse direction, at each phoneme change the words of the lexicon are updated, as are the phonotactic statistics. Consider the hypothetical lexicon fragment in Table 3.

Table 3 approx here

The four labeled columns of Table 3 indicate the four transitional probabilities that are calculated in the model: counts for each of the segments that occurs immediately after the phoneme of interest (..k^j__..), two segments after (..k^j__ __..), immediately before (..__k^j..), and two segments before (..__ __k^j..). It is assumed that both longer range, as well as higher order, dependencies are less relevant and less well correlated to typological patterns (although certain harmony, and anti-harmony (OCP) patterns may be exceptions to this). The phonotactic measure used in the model was a predictability score for segment *x* given that /k^j/ was present in relative position *w* (Maddieson & Precoda, 1992). See the Appendix for details. The phonotactic grammar can be expressed as a single number, an overall *Phonotactic Expectation* score for the likelihood of the segment of interest occurring in a given word:

(6) Phonotactic Expectation: The sum of the predictability for k/k^j occurring
(a) with the following segment,
(b) the preceding segment,
(c) the segment two segments subsequent,
(d) and the segment two segments prior.

As can be seen from Table 3, the number of /k^jɪ/ sequences is relatively large. This should not be surprising given the phoneme changes of Table 1. There are also a large number of /k^j/ phonemes that appear at the beginning of words; this results from the syllable structure of the model lexicon in which all syllables have onsets. Note that there are also a large number of words in which the consonant /p/ occurs two segments after /k^j/. The syllable structure again plays a role in this development. All /k^j/'s (except word-final ones) developed from /k/'s in the environment of a following vowel, therefore all these /k^j/'s will appear two segments prior to a consonant: .. k^jVC.. However, the consonant in that position can be any of the set {b,p,g,k,d,t,m,n,z,s}. The fact that /p/ appeared more frequently than any other consonant in this position (for this particular lexicon) must be attributed to the fact that it happens to occur quite frequently immediately after /ɪ/ (see the /k^jɪ/ column of Table 3, where 6 out of 10 words also contain the sequence /k^jɪp/). The ..k^j__ p.. phonotactic is strengthened by /p/ occurring also after other vowels. Comparing columns 3 and 4 of Table 3, we see that this sequence occurs in 10 out of 17 words. This result is completely serendipitous, and may be due in large part to the generally higher marginal probability of /p/ in this lexicon. Once the association has been established – a phonotactic expectation greater than zero – it continues to reinforce itself by boosting the probability of subsequent palatalizing changes in words with the same sub-structure.

The attributes of individual lexicons emerge through a random sampling process from an underlying probability distribution over segments, syllables, and word lengths. In general, 'unnatural' phonotactics like ..k^j__ p.. can arise under circumstances in which there is either non-uniformity in the underlying distributions, or in the sampling process. (See the Appendix for the specifics of lexicon construction.) The outcome of the simulations described below will provide an estimate of the frequency with which such 'unnatural' grammars occur given the current model.

Representations

A large body of work has shown that stored linguistic representations exist in a web of numerous associations. That acoustic, phonetic, semantic, morphological, social, and other properties are encoded by speakers has been shown in a variety of experimental paradigms (see, among many others, Dell, 1986; Lupker, 1979; Newman, Sawusch & Luce, 1997; Rosinski, 1977; Vitevitch, Luce, Pisoni & Auer, 1999). However, it will turn out to be the case that even a two-dimensional categorization problem is complex enough, both computationally and linguistically, to illustrate a number of fundamental theoretical issues. For this reason, as well as simplicity and clarity, such a reduced phonetic space will be used for modeling purposes (see Figure 2). Each phoneme category will be represented by a cloud of tokens in this space, where each token appears in a particular spoken word to which the listener has been exposed (Bybee, 2001; Goldinger, 1996; Johnson, 1997; Pierrehumbert, 2002, 2003; Pisoni, 1973).

The first two formants are generally taken to be adequate for capturing most vowel distinctions (e.g., Peterson & Barney, 1952). The acoustic distinction between plain versus palatalized segments may be best described in terms of other properties, such as peak spectral frequency or duration of aperiodic noise (see Guion, 1998). However, as the source of the phoneme split is taken to be the degree of frontness/height of neighboring vowels, the first two formants of the vowel following the k/k^j segment will determine the phonetic location of that token.

Figure 2 approx here*

Recall that, by hypothesis, final jer deletion (e.g., /dakɪ/ > /dak^j/) is responsible for initiating the novel /k^j/ category (indicated by the large white circle in Fig. 2). Since some phonetic position for those tokens must be established, word-final /k^j/ tokens will be placed in the phonetic location of the historically following /ɪ/ vowel. The black circles in Fig. 2, representing tokens of the /k/ phoneme, are distributed throughout the space, since there is nothing, *a priori*, that limits the distribution of vowels following /k/ (it is a fully contrastive segment). This distribution, at least initially, recapitulates a two-dimensional vowel space plot. The phonetic-space representation is also isomorphic at this time with the phonological distribution of the phones (i.e., pre-vocalic occurrence). Neither of these two properties are guaranteed to hold at later model stages as the phoneme changes continue to apply.

The Learning Mechanism

The learning task in this model is the mapping of phonemic representation to phonetic word token. More specifically, the learner must decide whether a given phone should be categorized as a /k/ or a /k^j/. This builds on previous work on phone categorization within an exemplar framework, beginning with analogical attraction (Blevins & Wedel, 2009; Pierrehumbert, 2001, 2003; Wedel, 2006, 2007). There is considerable evidence that analogy, or similarity, plays a role in categorization and speech processing (see, among others, Bailey & Hahn, 2001; Frisch, 1996; Goldinger, 1996; Johnson, 1997; Stemberger, 1990). For categorization purposes, the relative similarity of the current token to each of the competing categories must be computed.

In the exemplar literature the similarity function often involves a sum over the acoustic distance between the token to be categorized and each member of a candidate category (Luce & Galanter, 1963; Nosofsky, 1988; Pierrehumbert, 2001). The competition in the present instance, however, involves a non-typical linguistic context: category split within completely overlapped distributions. This type of categorization problem is known to be computationally difficult (see proposals for learning non-separable data (Burges, 1998; Vapnik, 1995), and Goldwater's (2007) model for unsupervised non-parametric learning of morphological categories via hyper-parameter setting). In the case of /k^j/ formation, not only is the new category a proper subset of the parent category, but the parent category is also orders of magnitude larger. Furthermore, as is illustrated by Fig. 2, the location of the /k^j/ category center is well within the /k/ category; which is to say that at the time of the creation of the /k^j/ category, there are a number of existing /k/ tokens which are produced with a more 'palatalized' articulation (higher F2, lower F1).

For these reasons, it can be shown that summing over all category tokens will prevent a new phoneme category from ever arising. The /k/ category will immediately reabsorb any such newly created /k^j/ tokens. Restricting the computation to a smaller window (manipulating Nosofsky (1988)'s similarity scaling term) will not solve the problem; windowing, in fact, actually worsens the situation, since it excludes consideration of differences in breadth of distribution. Computing over the token means for each category, however, will allow separation to be achieved in some cases. This works precisely because the /k/ category is much larger, and much more widely distributed.

An average /k/ location will, typically, be located somewhere towards the center of the (F1, F2) space, whereas the average over the few /k^j/ tokens will center around the canonical /ɪ/ position. These are essentially prototype locations, which have been argued for as alternatives to a homogenous exemplar space (e.g., Oden & Massaro, 1978; Rosch, 1977; Samuel 1982). To avoid category collapse, we will use an average over tokens to compute a single distance. However, the final metric will be a cross between a pure prototype and a pure exemplar computation. Under the assumption that larger categories possess an advantage due to frequency priming, the similarity function will be scaled by the token sizes of the categories (see Appendix for full mathematical details).

The palatalization strength of the vowel acts independently of the phonetic location of the /k^j/ category center – it depends only on the phonetic location of the immediate vowel context. This term captures the known facts about perceptual confusions (see Guion, 1998), which are assumed to be active synchronically as well as diachronically in the types of phoneme changes shown in Stage 2 of Table 1.

There is a third factor affecting categorization in this model: the listener/learner's expectation that a token of a given category will occur in the observed phonological environment (see discussion under Grammar). There is considerable evidence that phonotactic regularities affect speech processing (see, for example, Albright & Hayes, 2003; Kabak & Idsardi, 2003; Moreton, 1999; Pitt, 1998; Vitevitch et al., 1999). The effect of prior language knowledge on sound change, however, is an open question.

Including a top-down constraint on natural phoneme change allows for unnatural associations to develop in the simulations described here. It also allows, in theory, for changes to be blocked just in case they disrupt an existing regular structure. This is a long-cherished generativist postulate. However, the disruption of ordered grammars – resulting in either opacity or unpredictability – has been frequently noted in the historical record. Thus, the prevailing view holds that diachronic change is largely 'blind' to its effects, constrained only by phonetic naturalness. For this reason I attribute only a small relative force to a top-down constraint on phoneme change. This is also consistent with experimental work which, although it shows significant differences in perceptual judgments depending on the speaker's native language phonotactics, also tends to find the absolute size of those differences to be rather small, even for the case of inviolable phonotactic restrictions (e.g., Dupoux, Kakehi, Hirose, Pallier & Mehler, 1999). Similarity is therefore weighted much more heavily than Phonotactic Expectation for categorization probability in the model (see Appendix).

The final force that affects categorization comes from production rather than perception. Periodically, all stored tokens are replaced with new values in phonetic space. The locations of the new tokens are shifted in the direction of a local average. This reduces variance, consolidating the phonetic position of the category as a whole. While the boundaries of categories are continuously changing as tokens are added and removed, this homogenizing effect acts to keep categories from spreading out too far, or losing continuity. This is one way to model entrenchment, a phenomenon by which the range of variants becomes narrower, rather than broader, over time, due to motor tuning, or articulatory practice (see Pierrehumbert, 2001). The characteristics of the learner are summarized in (7).

(7) The Learner: categorizes phones within a F1-F2 phonetic space as a function of
- Phonetic similarity to existing phoneme categories
- Degree of phonetic palatalization
- Phonotactic expectation of phoneme within word

SIMULATIONS

The full model implementation is a balance between simplifying in order to make an extremely complex system tractable, and adding complexity in order to make underspecified theories testable. A given language consists only of a lexicon and a set of phonotactic statistics; segment tokens occupy just two dimensions of phonetic space, and communication consists of a single speaker/listener feedback loop. However, segments, and changes to segments, occur within a naturalistic lexicon, prior language knowledge is a factor in the model, and phonetic similarity is explicitly defined and implemented. In the simulations that follow the set of output grammars will be described as a function of initial lexical conditions. We will identify which model features are responsible for which facets of this distribution, and we will demonstrate the sequence of events necessary to produce particular pathological outcomes.

For a single iteration of the model, a lexicon is randomly generated. The segment tokens of this lexicon are then subjected to a sequence of phoneme changes; this comprises a model Event. The category centers (in F1-F2 space) and the phonotactic expectations are updated after each individual phoneme change. At Event 1 all changes in Table 1 apply. Subsequent Events involve re-application of just the Stage 2 phoneme change. Entrenchment, in which a given category is consolidated by shifting each token towards the phonetic mean of its neighbors, operates after every 3 Events (see Appendix for details).

Figure 3 shows the evolution of the phonetic tokens of the two categories of interest over a simulation involving 50 Events. In panel 1, the /k^j/ category (white circles) is instantiated in the upper left quadrant: the high, front vowel environments. The category expands due to re-categorization – Stage 2 changes that add tokens before /i/, /ɪ/, and /e/. However, fortuitous statistical regularities within the smaller /k^j/ category add phonetically unnatural palatalizing contexts (i.e., o,u,ʊ,a). This slows the movement of the /k^j/ category center towards the highest, frontest position.[5]

[5] When non-natural contexts are added, e.g., ka>k^ja, the phonetic location of the palatalized variant is taken as the center of the existing k^j category with the addition of a random perturbation. The perturbation is half the size of that used to generate the variability for the natural palatalizing contexts in order to prevent excessive category spread. The addition of palatals in non-natural environments does not push the phonetic category center towards those vowels per se, but it reinforces the category at its present location. This effectively introduces resistance to the natural high, front drift.

The size of the /k/ category steadily decreases as more /k^j/ tokens are created. However, it remains at all times the larger category. For tokens equidistant to the two category centers, the /k/ category will receive a higher similarity score due to its larger size (see Appendix for details of the similarity function). This will also be true of tokens which are closer to the /k^j/ category, but not far enough away from the /k/ category to escape its greater influence. Due to the original lack of phonetic separability of the two categories, many tokens that fall close to the /k^j/ prototype also fall close to the /k/ prototype. For tokens in this region, the high, front vowel palatalizing bias can be overcome by the slightly higher degree of similarity to the unpalatalized category. Thus the two exemplar clouds drift apart and back together, remaining partially overlapped throughout.

Figure 3 approx here.*

The relationship between the plain and palatalized velar phoneme can also be described in terms of a phonological distribution. See Figure 4. The /k^j/ category is initiated before an historic /ɪ/, and spreads to other high, front vowel environments. At the same time, non-palatalizing vowel environments are being added, although this proceeds more slowly. The /k^j/ category steadily expands until, by Event 50, /k^j/ is found in all possible pre-vocalic environments.

Figure 4 approx here

In a strict sense, the distribution of /k^j/ is unpredictable, and thus it is in a contrastive relationship with /k/. This distribution, however, is highly skewed, and the perceptual bias for /k^j/ in high, front environments is readily apparent from the first to the final panel of Fig 4. It may be more accurate to describe such a distribution as a 'semi-allophonic' relationship, where /k^j/ is only a 'partial', or 'quasi', phoneme (see Hall (2009), Hume and Johnson (2003), and Scobbie & Stuart-Smith (2008) for discussion of gradient treatments of phonological relationships). This fuzziness is inevitable in describing the outcomes of the model, implementing, as it does, stochastic processes over individual tokens. In fact, this kind of fuzziness may be inevitable when dealing with non-idealized natural language data. For present purposes, however, we will be content with labeling this the 'contrastive' outcome: one of four logically possible grammatical relationships between the /k^j/ and /k/ phonemes, the other three being allophonic, neutralizing, and anti-allophonic. These, and the conditions for which they are predicted under the current model, will be discussed below. Before doing so, however, we will examine the behavior of the model in more general terms.

General Model Behavior

The inherent asymmetry in misperception, such that k>k^j is more likely than k^j>k, ensures that the /k^j/ category will grow steadily in the upper left quadrant of formant space. The dissimilatory phoneme change (Stage 3 in Table 1) then introduces /k^jo/ tokens into the distribution. However, the only way for the palatal variant to spread to the other non-palatalizing environments (ʊ, u, and a) is via pressure from the phonotactic grammar. Furthermore, since the Stage 3 change applies only to a subset of /k^jek/ sequences, this typically results in just a handful of new /k^j/

tokens before /o/. Thus, the majority of /ko/ > /k^{j}o/ changes are also due to the phonotactic grammar.

'Unnatural' sequences involving /k^j/ in the context of non-palatalizing vowels come about indirectly due to correlations that develop between /k^j/ and segments at one remove. Recall that in the example from Table 2 a random and phonetically unmotivated correlation arose between k^j and p separated by a vowel. Originally, this association was due entirely to "natural" words, such as {k^{j}ɪpsɪs, k^{j}ɪpkon, k^{j}ɪpset, k^{j}epʊssu}, in which the following vowel was the actual source of the phoneme change (Stage 2, Table 1). Once the k^j__ p phonotactic was established, however, it acted to promote phoneme change in other words, regardless of the vowel context. This led to such tokens as k^{j}ʊpʊs and k^{j}ʊpʊssʊt. And these additional tokens, of course, only acted to increase the strength of the phonotactic.

Random sampling is much more likely to produce spurious correlations in small samples than in larger ones.[6] Thus, phonotactics that arise within the initially very small /k^j/ category will tend to be stronger than those among the much more numerous /k/ words, generating a net bias in the /k/ > /k^j/ direction. Though the strength of those phonotactics may decrease as the /k^j/ category grows, the directionality of the bias is not expected to change as a significant size disparity remains between the two categories.

Phonotactic forces act to expand the phonological distribution and homogenize the phonetic distribution of the /k^j/ category. The entrenchment/production mechanism, in contrast,

[6] Larger samples from some underlying population are more likely to be representative of that population than are smaller ones, if both are produced randomly. The segments in each word which do not participate in the phoneme change are effectively randomly sampled from all segments that occur in that relative position. For a uniform underlying distribution, under-sampling should produce some instances of non-uniformity. The underlying distribution of segments in the model, however, is not uniform, but Zipfian, with certain segments being much more likely over-all than others (see Appendix). Under-sampling in this case can act to flatten the distribution, but it can also act to reinforce even more an already skewed distribution.

acts to concentrate the phonetic distribution. For early Events, entrenchment reinforces the location of the /k^j/ category in the upper left quadrant of F1-F2 space. In turn, this increases the likelihood of phonetically nearby (k / __ {i, ɪ, e}) tokens being re-categorized as /k^j/'s. If entrenchment proceeds too quickly, however, it can actually act to exclude phonetically palatalized variants from the /k^j/ category. There is, essentially, a competition between spread and consolidation. If entrenchment progresses too quickly – before enough /ki/ tokens have been re-categorized as /k^ji/'s through natural assimilation changes – such tokens will be increasingly less likely to shift as the /k^j/ category center becomes more narrowly focused in a lower, less front region of phonetic space (centered around the original /ɪ/ context for palatals). If entrenchment progresses too slowly, then the category spreads out, becoming less differentiable from its parent /k/ category, and exerting a weaker pull on tokens at its peripheries.

Initial Conditions

Due to factors in the design of the model the /k^j/ category is pre-determined to evolve in certain ways. All simulations will demonstrate initial rapid expansion to phonetically nearby tokens, accompanied by phonological expansion to all natural pre-vocalic contexts. In most cases, a slower, but continuous, addition of non-natural phonological contexts will follow, preventing the phonetic distribution from drifting to an extreme. Even with this being the case, it is not true that all iterations of the model will produce the same results. The particular grammar that develops will depend strongly on the particular lexicon with which the simulation is run.

The lexicon of Figs. 3 and 4 – call it Lexicon A – is characterized by large numbers of ki and kɪ tokens, as well as large numbers of kV tokens in general. These initial conditions produce a particular type of phonological distribution, one that can be termed contrastive, or semi-contrastive. The two velar variants are not predictable from context, but neither are they completely homogenous in their pre-vocalic distributions – they show a persistent 'naturalness' bias. This is accompanied by phonetic ambiguity in the productions of certain tokens of each category.

Lexicon B, on the other hand, has been selected to represent a situation in which /k/ occurs initially in low numbers before /i/ and /ɪ/, and high numbers before other vowels. These initial conditions are apparent in the much denser distribution of black circles on the right, than the left, side of the second panel at the top of Figure 5. The relevant difference to Lexicon A is that the two category centers begin as fairly separated. Initial separation is then reinforced over time with both Similarity and Entrenchment forcing the categories further and further apart. The phonetics interact with the phonotactics to produce the final phonological distribution in the second panel at the bottom of Fig. 5. Although this outcome appears quite similar to the outcome for Lexicon A, there are both fewer palatal variants in non-natural vowel contexts, and fewer plain variants in naturally palatalizing contexts. The distribution that is represented by this lexicon could be described as a contrast which is effectively neutralized to the /k^j/ variant before high, front vowels.

Finally, Lexicon C is characterized by low counts of /k/ before the non-palatalizing vowels. In fact, there are no /ko/ or /kʊ/ tokens at all in this particular lexicon. The low numbers of non-palatalizing contexts act to preserve the naturalness of the /k^j/ category. Even if the necessary phonotactics arise, the dearth of unnatural candidates makes the addition of unnatural contexts less likely. This initial disadvantage is reinforced quickly, as large numbers of natural tokens are rapidly absorbed into the /k^j/ category without any counter-balancing pressures.

Phonotactics reinforcing the perceptual bias only grow stronger, washing out any weaker associations. For Lexicon C, the final outcome largely preserves the allophonic relationship: /k^j/ only occurs before high, front vowels. In this particular simulation, there remains a certain amount of acoustic overlap between the two categories, with an ambiguous region in which tokens may continually shift back and forth. Figure 5 show the three lexicons side by side; the top row gives the phonetic distribution at the beginning of the simulations, and the bottom row shows the phonological distribution by Event 50.

Figure 5 approx here

The final type of phonological distribution – an anti-allophonic distribution – is characterized by the opposite pattern to that of the Lexicon B results. /k/ appears only before high, front vowels, and /k^j/, only before the non-palatalizing vowels. However, it can be shown that this outcome will not occur for any type of lexicon. Since such a system is, to my knowledge, unattested, this represents a success, rather than a failure of the model. The anti-allophonic distribution is also an instance of a larger class of implication-violating phonotactic grammars. Part of the debate over Universal Grammar involves establishing what theoretical mechanisms are necessary to ban such 'anti-natural' grammars. In later sections we will turn this question around and ask instead under what specific conditions such grammars could arise at all. Before turning to that discussion, however, let us consider a somewhat less pathological outcome that is predicted by the current model.

Unnatural Phonotactics

As illustrated previously, the fact that non-phonetically based phonotactics can be learned is the factor that allows for the k^j allophone to achieve contrastive status under certain circumstances.

This happens indirectly via non-sequential correlations (e.g. k^j__ p) that add non-palatalizing vowel contexts to the distribution (e.g. /k^jap/). Under the assumption that secondarily articulated phonemes enter inventories in this way, and further, that they achieve fully contrastive status in at least some languages, this is a desirable feature of the model. However, this may be considered an undesirable property of a synchronic grammar. If such phonotactics are not common in attested languages, then the model may over-predict their occurrence. To answer this question we will have to decide how large an association must be before it achieves a recognizable synchronic status (i.e., is recorded by linguists). In the meantime, we will estimate the predicted rate by comparing the strengths of natural and unnatural phonotactics for simulations run over 1000 different lexicons.[7]

Recall that there were four bigram statistics calculated for the phonotactic grammar. Each panel of Figure 6 corresponds to one of the four phonotactic measures for /k^j/, with respect to each of the other segments in the inventory (the x-axis). The y-axis indicates the number of languages (out of 1000 total) that reached a critical threshold of predictability for each of those segmental contexts. That threshold was set at 60% of the strongest phonotactic in that particular language. For example, for one language, the highest predictability phonotactic was for the /k^ji/ sequence at .33. Any additional associations with predictability scores larger than .20 in that language contribute to the count in Fig. 6.

Figure 6 approx here

---

[7] For 1000 different randomly produced lexicons, the phonotactic grammar was assessed after 6 Events. This number of Events was chosen to reduce computation time, but still proved sufficient for the majority of the /k^j/ tokens to be generated, and the general behavior of the system to be clear.

Most associations are near zero. Panel 3 clearly shows the Natural associations that are the source of the phoneme change: high front vowels. There is also a relatively high rate of /1k^j/ association (1 = the beginning of the word), as seen in panel 1. That /k^j/'s are likely to occur at the beginnings of words is due to the syllable structure of the language. All one-syllable words, for example, take either the form CV, or CVC. Any /k^j/'s that appear in one-syllable words must also appear at the beginning of the word.

Additionally, out of 1000 separate simulations, there is one instance each of a threshold-reaching phonotactic between /s/, /d/ and /t/ in the position immediately preceding /k^j/. Syllable structure is also responsible for a 100% correlation between /k^j/ and a consonantal segment two positions later. The unnatural associations arising in this position are shown in panel 4 of Fig. 6. Individual lexical idiosyncrasies account for high predictability scores between /k^j/ and any particular consonant for any particular language.

There are 219 instances of unnatural associations over the set of simulations reported here. Although multiple instances occasionally arise within a single language, the vast majority of lexicons contain a maximum of 1 unnatural association. Thus, unnatural associations can be said to arise in roughly 20% of all lexicons. The height of the points in panel 4 of Fig. 6 indicate that the associations are fairly evenly distributed among the possible consonantal segments, with slightly higher probability of a given lexicon containing a significant k^j__b association than a k^j__ p association, for example. 3 out of the 219 total unnatural associations occur in isolation, that is, without any other phonotactics reaching a significant level (see 7a for an example); an additional 39 meet or exceed the predictability strength of the co-occurring natural phonotactics within the same language (see example in 7b); and the remainder – 177 instances – are weaker than the natural phonotactic associations within the language (7c).

(7a) Lexicon 724:
Pr(k^j |_n) = 0.496 e.g. */kʲɪnʊ/*

(7b) Lexicon 776:
Pr(k^j|ɪ) = 0.242 e.g. */kʲɪtak/*
Pr(k^j|i) = 0.269 e.g. */sumkʲig/*
Pr(k^j|_g) = 0.321 e.g. */sumkʲig/*

(7c) Lexicon 804:
Pr(1|k^j) = 0.164 e.g. */kʲɪk/*
Pr(k^j|ɪ) = 0.256 e.g. */kʲɪk/*
Pr(k^j|e) = 0.266 e.g. */takʲepi/*
Pr(k^j|_k) = 0.185 e.g. */kʲɪk/*

Examining the lexicons after a certain number of Events and assigning a significance threshold, as has been done here, gives an approximation of how many of those phonotactics can be expected to survive within a given synchronic grammar, producing 'crazy rules' (Bach & Harms, 1972). Clearly, there are a non-zero number of these, although whether the model predicts more unnatural associations than are found in nature is difficult to determine. Of the 219 associations, some may be objectively too small to be learned, or robustly measured, or they may simply have been over-looked in typological reporting. The best candidates for discovery are the lone associations, and those that actually exceed the expected natural associations, leaving us with a conservative estimate of 39 out of 1000 cases of over-generation.

PHONOLOGICAL UNIVERSALS

At the beginning of this paper reference was made to an implicational hierarchy for velar palatalization. Implicational hierarchies (especially as invoked in classical Optimality Theory) are typically couched in categorical terms, taken to require that a contrast be instantiated, or completely neutralized, that a sequence be completely banned, or freely surfacing. The relation i > ɪ embodies the hypothesis that vowel i is universally preferred over vowel ɪ as an environment for palatalization. If it is found in a particular language that velars palatalize in the context of ɪ, then the theory predicts that velars will also be found to palatalize in the context of i. This also implies that velars should palatalize in the context of i diachronically before they palatalize before ɪ (if at all). A violation of the universal would be entailed if any of the following were true in a given language: synchronic alternations in which plain velars palatalized only in the context of ɪ, and not i; a diachronic change that shifted plain velars to palatals only in the context of ɪ and not i; a lexical distribution in which palatalized velars could be found in the context of ɪ, but not in the context of i.

In each of the implication-violating scenarios above, it is implied that i can never be the context for the palatal variant. This type of categorical outcome does not occur, however, in the stochastic model just described. The grammar is a set of gradient phonotactics, and the segments themselves exist along a continuum of contrastiveness. Rather trivially, we could state that our model never violates the implicational hierarchy because it fails to treat the vowel types as monolithic categories – there will always be some palatals occurring before each of the palatalizing vowels. It may be more instructive, however, to formulate the substance of the universal in the following way: vowel ɪ ought never to occur *more frequently* in palatal contexts than does vowel i. When the relative number of k^jɪ bigrams is greater than the relative number of k^ji bigrams, or the relative number of k^jo bigrams is greater than the relative number of k^je bigrams, etc., a violation of the proposed universal palatalization hierarchy is counted.

The actual measure we will use to calculate a gradient violation of the hierarchy is a ratio of k^jV sequences to kV for each vowel context: how much more often the vowel occurs with the palatal than the plain velar.[8] Numbers above .5 indicate a palatalizing preference, while numbers below .5 indicate a palatalizing dispreference. Figure 7 shows the average over all 1000 lexicons: a monotonically decreasing function as one moves rightward in the hierarchy of (3). This is exactly what should occur according to the predictions of the gradient implication; *on average*, the model produces no distributions violating palatalization universals.

Figure 7 approx here

Gradient implicational violations within particular lexicons are another matter. Figure 8 gives the number and degree of implicational violations by language. Just in case the ratio measure is greater for the vowel neighbor more rightward on the implicational scale, the degree is calculated by taking the difference between the two vowels. The total number of such violations (out of a possible total of 1000) is printed next to the boxplot for each vowel pair. Fig. 8 shows that there are recurring violations between the i-ɪ pair, as well as between the e-o pair. All other pairs show negligible numbers of violations, and of small degree. The relatively large degree of the e-o violation can be explained by the inclusion in the model of a phoneme change which converts certain /k^je/ sequences to /k^jo/ (Stage 3 in Table 1). Violations are the most common for i-ɪ sequences (4.7%) due, in part, to the fact that /ɪ/ possesses a similarity advantage at the inception of the category (due to word-final /k^j/'s at the phonetic location of the historically following /ɪ/).

---

8 Phonotactic predictability (6) was measured using a relative transitional probability: the probability of the vowel appearing in the context of a preceding k^j, versus the probability of the vowel in any context. Calculating the conditional probability of the vowel given the palatal in this way makes the statistic a measure of how well the given vowel predicts k^j – relative to how well other vowels do. This is, of course, not the only way a predictability measure could be formulated. One could ask, instead, what the probability of k^j was, given the vowel: a measure of how often the vowel was associated with k^j rather than with other segments. This measure, in fact, may be more appropriate to the assessment of implicational violations, especially when restricted to a comparison between k^j and k. Thus, the proportion of k^jV sequences to (k^jV+kV) sequences is the measure used in Fig. 7. The same measure is used in Fig. 8, where implicational violations are assessed on a language by language basis.

*** Figure 8 approx here***

According to the implicational rule as redefined, all languages shown in Fig. 8 are impossible, and represent a mismatch between the model predictions and typology – an over-generation rate of 9.2%. However, it should be noted that none of the lexicons in Fig. 8 represents a consistent reversal of the implicational scale. In actuality, the only reversals of relevance would be among the first four vowels, since they are the only ones which allow $k > k^j$ changes based purely on phonetics (rather than requiring phonotactic pressures). To qualify as a reversal then, we can relax the requirements such that the relative frequency of palatalized tokens decreases in the order {u,ʊ,a} > o > e > ɪ > i. There are no lexicons like this resulting from the above simulations. In fact, there are no lexicons that contain more than one violation; each data point in Fig. 8 represents a unique lexicon.

Reversed Implicational Scales

In this section we will show what general types of phoneme change are necessary to produce a truly reversed implication hierarchy, and in the next, we will attempt to construct a possible historic trajectory for such an outcome. These two sections will demonstrate that the *lack* of pathological patterns is a result of principled constraints on the Source component, and does not come about from over-simplifying the model, nor as an artifact of specific parameter choices.

Table 4 depicts the three broad grammar types classified as possible outputs of the model: Column 1: Implication Compliant, Column 2: Gradient Implication Violating, and Column 3: Implication Reversing. The "Implicational Hierarchy" column provides a schematic exemplifying the relative proportions by vowel (in implicational order along the x-axis) that captures each of the types. Initially, all lexicons reflect Implication-Compliant grammars (vacuously, since there are no palatalized segments). The bias that produces higher rates of assimilatory change before higher, fronter vowels will act to preserve the implication. See Column 1.

Table 4 approx here

Gradient violations can arise due to dissimilatory phoneme changes that reduce the k^jV bigram counts for a given V to the same degree that they increase the bigram counts for any vowel rightward in the hierarchy. This is illustrated by the change k^ji > k^jɪ, a documented change within the Slavic family. If this change had been included in our original Source model it would have acted to increase the number of i-ɪ implicational violations. See Column 2.

Another way to produce a statistical violation is by an increase in the kV bigrams involving any vowel that is not rightmost in the hierarchy. An influx of new words, of which a large enough number contain /kɪ/ tokens will reduce the proportional measure for /ɪ/, all else being equal. Enough new /kɪ/ tokens can cause a implication violation at ɪ-e. Such a wholesale increase in tokens could potentially be the result of a large borrowing from a neighboring dialect. See Column 2.

To reverse the implicational hierarchy wholesale requires a qualitatively different type of change. Consider a hypothetical change such as that given in (8). Such a shift would lower the bigram k^ji counts, while raising the bigram k^ja counts, thus reversing the natural relative implication degree between /i/ and /a/. The natural bias that exists in the Source, however, would restore many of the k^ji bigrams. The hypothetical change in (9) would act to increase the number of /ki/ bigrams, a complementary change that also acts to lower the implication degree of /i/, while raising that of /a/. (8) and (9) in combination would need to be stronger than the natural bias for palatalizing in the context of /i/, such that /a/ not only achieved a higher implication score, but higher to a degree than allowed /i/ to be characterized as an unlikely palatalizing context.

(8) tok^ji > tok^ja

(9) teka > teki

In fact, the implication score for /i/ must be the lowest of all the vowels, situating it at the very end of the continuum. /ɪ/, in turn, must be of the second lowest implication degree, and /e/ of the third, just as /u/ must be of the second highest implication degree. For a full implication reversal, all vowels must effectively swap places with their implicational complements. See Column 3. However, even swapping vowel contexts directly, as in (8) and (9), cannot guarantee this result. Even for the relatively simple model of this paper it is not possible to fix strict conditions for the final distribution without strict requirements on the input lexicon. Structure is more likely to be disrupted than it is to be spontaneously generated. This is even more true of sequences of indirect changes – ones which are better motivated than the context-free changes of (8) and (9). For a stochastic, context-sensitive, outcome-blind model, there exists an extremely narrow range of diachronic paths leading to the reversed implication grammar (see the Appendix for more concrete illustrations of this argument). It is concluded that it is thus quite improbable that such a grammar could result by chance. Instead of imposing external constraints on the model to avoid these outcomes, we must resort to contortions in order to produce them at all.

FINDINGS & CONCLUSIONS

Computational modeling forces us to select particular hypotheses and then test them in such a way that the theory as a whole produces an actual output (a fully specified lexicon). By this method inconsistencies or gaps in existing theories are often revealed, as well as incompatibilities between theories pertaining to different sub-domains of linguistics. For example, one foundational premise of this paper – that new phoneme categories arise through the erosion of existing ones – required similarity metrics with an atypical dependence on category size (see the discussion of the Learning Mechanism). Without this modification those new categories would be immediately reabsorbed into the overwhelmingly larger original category. If one accepts this premise, then such a modification is required. If such a modification is rejected on independent grounds, then some other mechanism for phoneme genesis must be proposed.

In a similar vein, traditional generative theory tends to render universal linguistic tendencies as absolutes. This allows for clear theoretical predictions, but it makes it harder to compare theoretical grammars to actual language data, as well as to the results of models which operate over words rather than a space of grammars, rules, or constraints. Requiring phoneme change to apply to lexical items in a stochastic way forces a reconsideration of what the linguistic input really consists of. Requiring learning to take place over those same lexical items introduces a source of random variation to the predicted distribution of phonological grammars. If this range of variation is broader, or narrower, than that of the actual typology in some way then the model must be modified, otherwise the model stands as an instantiation of a sufficient theory of the phenomenon.

The possible grammars predicted by the model of this paper include Natural ones – a certain degree of allophony, or neutralization, between /k/ and /k^j/ – , Contrastive ones – /k/ and /k^j/ largely unpredictable from context –, Unnatural ones – /k^j/ occurrence becomes strongly correlated with a random segment – , and Partial Violations – occasional numerical deviations from the palatalization hierarchy of (3) – . Unnatural phonotactics occur roughly a third of the time, although they most often occur within a system that is Natural on the whole, with only 42 out of a possible 1000 languages being unambiguously 'Unnatural'. Numerical implicational violations are rather more common (92 instances in 1000 languages), but they too appear within an otherwise largely Natural system, and no single lexicon contains more than one such pairwise reversal. That is, the model produces no 'anti-natural' grammars, grammars which would systematically violate the implication direction.

I take the results of these simulations to indicate that the current model is sufficient to account for the typological facts as far as they are known. This conclusion predicts that typological reporting is somewhat incomplete: either missing data or missing languages that could comprise a 4.2% rate for Unnatural patterns, and a 9.2% rate for singleton gradient implication violations. See Evans and Levinson (2009) for estimates of current typological coverage, and Mielke (2008) for data on the frequency of phonological patterns conditioned by non-natural phoneme classes. It is thus concluded that there are no additional components – in particular, a substantive restriction derived from Universal Grammar – that must be added to the model in Fig. 1 in order to produce the correct distribution. The model fails to over-generate rare, or unattested, – and theoretically undesirable – grammars (de Lacy, 2006; Kingston & de Lacy, 2006; Kiparsky, 2006, 2008). In fact, the results suggest a reversal of the premise. Rather than asking what mechanisms are required to avoid this outcome, the question ought perhaps to be what conditions are required to achieve it.

More generally, the results of this paper act as an existence proof for a model that can produce a relatively narrow typology without placing explicit requirements on the end-state grammars themselves. The claim, however, is not that such constraints do not exist in human learners, only that they are not theoretically necessary. In fact, Wilson (2006) finds that learners in a language game experiment show an asymmetry in which they are more likely to palatalize before /i/, having heard the pair gefə-dʒefə, than they are to palatalize before /e/, having heard the pair gibə-dʒibə. He attributes this result to a substantive bias in accord with the implicational hierarchy of palatalization – rather than a categorical substantive constraint. Such a bias may be innate, or it may arise from our phonetic experience with our native language. I do not refute these findings. The present work argues that, although human learners may have such a bias, it is not necessary to generate the correct typology.

It is true, of course, that the model predictions are only as representative as the values of the various parameters that were used to produce them. If the set of lexicons, for example, turns out to be non-representative of natural language in some way, then the model results will be skewed. Similarly, if the Source is incompletely described then we may worry that the model is missing exactly the type of phoneme change that will act to consistently produce pathological grammars. There can be no doubt that, with only three phoneme changes, the model is missing many possible changes involving palatalized consonants. This only becomes a potential concern if any one of those phoneme changes represents a distinct *type* – one that will affect the lexicon in qualitatively different ways than the other types of phoneme changes already included. A considerable amount of analysis has gone into an attempt to be exhaustive in this respect, see especially the previous section and Appendix. But there is no guarantee that this attempt has been completely successful.

Of particular concern may be the drastically simplified representation of phonetic space used in this model. It has been shown that the degree and extent of sound shifts can depend on very fine-grained details of the sound context (see, e.g., Yeager et al., 1972). Additionally, acoustic measures such as peak spectral frequency and duration of aperiodic noise are known to be good indicators of the presence or absence of palatalization (see Guion, 1998). These parameters should therefore be included in a more complete model. However, I maintain that the addition of acoustic variables will not fundamentally alter the main result of this paper. To see this, let us first consider how the simple two-dimensional space affects model outcomes.

In fact, using pre-vocalic phonetic context alone captures an important typological finding. It has been noted that even for languages that are conventionally described as exhibiting contrast between plain and palatalized variants, an asymmetry in their distribution persists. This is true in Russian, for example, where palatalized consonants are more common before front vowels (see Kochetov, 2002). The stochastic nature of the model leads to outcomes that are gradient in exactly this way. Contrastive systems arising from a phoneme split are required to go through an earlier stage of allophony – conditioned only by following vowel. This diachronic source is reflected in the distribution of the two phonemes even after the categories have gone through multiple shifts.

The end result of adding additional acoustic variables will be that grammars reflect their historic origins still more closely. More parameters of variation mean more ways to differentiate word contexts from one another, thus limiting generalization. For example, when only the following vowel context is relevant, the words /dikip/ and /dʊkip/ are identical at the phonetic level. However, if preceding vowel context is also taken into account, then the two /k/ tokens will be shifted away from one another in phonetic space. This increases the chances that the two tokens will not share the same fate. Perhaps only the first word will trigger palatalization[9]. Similarly, the phonological distributions of the two phonemes will become more specific. Fortuitous phonotactics will be less likely to arise because more facets of the word context are implicated in the palatalization process, and thus the sequence of segments that would have to be shared by all words is longer. For example, whereas before an association between k^j and p could

[9] One might also decide to treat the palatalization shift itself as gradient, where not only the probability of palatalization, but also the degree to which a phoneme palatalizes, depend on phonetic context. This also acts to create more fine-grained distinctions between tokens. However, this option compicates the determination of category membership, and will, I suspect, tend to reduce the number of unique output grammars.

arise if there happened to be a lot of words with the sequence k^jip, now the association might require a lot of words with the sequence ik^jip. A diminished effect of unnatural phonotactics, in turn, will make undesirable grammars even less likely to arise than they were before. Thus the result that substantive constraints are unnecessary stands[10].

For the simple model of this paper, a two-dimensional phonetic space is specific enough to show the effects of random variation, and to demonstrate the importance of initial conditions on the evolution of the system. But it is also abstract enough to allow for generalization to occur in the form of fortuitous associations. A more complex model will need to meet those requirements as well.

The methodology in itself is flexible enough to allow model elements to be altered, added, or removed, thus permitting this initial attempt to be scaled up and improved. The proposed modifications need only be made concrete enough to implement. While a seemingly simple requirement, this paper has hopefully shown that it is by no means a straightforward task. The working hypotheses of the model are not merely implementational conveniences, but represent part of the theory of phoneme change. The theory is developed by revising those hypotheses. Individual components, or specific working assumptions, may need to be significantly altered to accommodate new findings, or new frameworks. The initial success of this model stands, however, in demonstrating the consequences of particular theoretical commitments, and in illuminating the true scope of the parameter space. Beyond even the particular phenomenon of this paper, the methodology itself remains as an important tool for testing and refining linguistic theory.

---

[10] In fact, one might now worry that a too high degree of specificity might preclude even contrastive systems from arising (as these require a certain level of unnatural phonotactic pressure). This raises a very fundamental theoretical question: how abstract linguistic representations actually are. One may be able to state the acoustic context to an arbitrary degree of precision, but that does not mean that all acoustic differences are encoded, or weighted identically in calculations of similarity, or for the purposes of determining statistical correlations. This suggests one direction for extending the work in this paper. Given a more realistic model of phonetic factors, how much filtering is required by the Learning Mechanism (i.e., how much of the phonetic information must be discarded) in order to produce a fully contrastive outcome.

References

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, *90*, 119-161.

Andersen, H. (1978). Perceptual and conceptual factors in abductive innovations. *Recent developments in historical phonology*, 1-22.

Aski, J. M. (2001). Prototype categorization and phonological split. *Diachronica*, *18*(2), 205-239.

Bach, E., & Harms, R. (1972). How do languages get crazy rules? In R. Stockwell & R. Macaulay (Eds.), *Linguistic change and generative theory* (pp. 1-21). Bloomington, IN:Indiana University Press.

Bailey, T., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, *44*, 568-591.

Beddor, P., Harnsberger, J., & Lindemann, S. (2002). Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics*, *30*(4), 591-627.

Bhat, D. (1978). A general study of palatalization. In J. Greenberg (Ed.), *Universals of human language* (Vol. 2, pp. 47-92). Stanford University Press.

Blevins, J. (2004). *Evolutionary phonology: the emergence of sound patterns*. New York, NY:Cambridge University Press.

Blevins, J., & Wedel, A. (2009). Inhibited sound change An evolutionary approach to lexical competition. *Diachronica*, *26*(2), 143-183.

de Boer, B. (2000). Emergence of vowel systems through self-organisation. *AI Communications*, *13*(1), 27-39.

Boersma, P., & Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, *32*(1), 45-86.

Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*, 121-167.

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

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*(03), 261-290.

Bybee, J. (2008). Mechanisms of change in grammaticization: the role of frequency. In B. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 602-623). Oxford, UK:Blackwell Publishing Ltd.

Chang, S., Plauche, M. C., & Ohala, J. (2001). Markedness and consonant confusion asymmetries. In E. Hume & K. Johnson (Eds.), *The role of speech perception in phonology* (p. 79-101). New York, NY:Academic Press.

Cutler, A. (1987). Speaking for listening. In A. Allport, D. McKay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 23-40). London:Academic Press.

Daland, R., Sims, A., & Pierrehumbert, J. (2007). Much ado about nothing: a social network model of Russian paradigmatic gaps. In *Proceedings of the 45th annual meeting of the association for computational linguistics*.

Dell, G. (1986). A spreading activation theory of retrieval in language production. *Psychological Review*, *93*, 282-321.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1568-1578.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429-492.

Fairbanks, G. (1965). *Historical phonology of Russian*. Poona:Deccan College.

Frisch, S. (1996). *Similarity and frequency in phonology*. Unpublished doctoral dissertation, Northwestern University, Evanston, Il.

Fujimura, O., Macchi, M., & Streeter, L. (1976). Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. *Language and Speech*, *21*, 337-346.

Gibson, E., & Wexler, K. (1994). Triggers. *Linguistic Inquiry*, *25*(3), 407-454.

Goldinger, S. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 1166-1183.

Goldwater, S. (2007). *Nonparametric Bayesian models of lexical acquisition*. Unpublished doctoral dissertation, Brown University.

Guion, S. (1998). The role of perception in the sound change of velar palatalization. *Phonetica*, *55*, 18-52.

Guy, G. R. (1980). Variation in the group and the individual: the case of final stop deletion. In W. Labov (Ed.), *Locating Language in Time and Space*, (pp. 1-36). New York, NY: Academic Press.

Guy, G. R. (2008). Variationist approaches to phonological change. In B. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 369-400). Oxford, UK:Blackwell Publishing Ltd.

Hajek, J. (1997). *Universals of sound change in nasalization*. Oxford, UK:Wiley-Blackwell.

Hall, K. C. (2009). *A probabilistic model of phonological relationships from contrast to allophony*. Unpublished doctoral dissertation, The Ohio State University.

Hamilton, W. (1980). *Introduction to Russian phonology and word structure*. Columbus, OH:Slavica Publishers, Inc.

Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, *39*(3), 379-440.

Hume, E., & Johnson, K. (2003). The impact of partial phonological contrast on speech perception. In *15th international congress of phonetic sciences*.

Hura, S., Lindblom, B., & Diehl, R. (1992). On the role of perception in shaping phonological assimilation rules. *Language and Speech*, *35* (1,2), 59-72.

Hutchins, E., & Hazlehurst, B. (1995). How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert & R. Conte (Eds.), *Artificial societies: the computer simulation of social life* (pp. 157-189). London, UK:UCL Press.

Hyman, L. (2002). Is there a right-to-left bias in vowel harmony? In *9th international phonology meeting*. Vienna.

Janda, R., & Joseph, B. (2003). Reconsidering the canons of sound-change: Towards a "big bang" theory. In B. Blake & K. Burridge (Eds.), *Selected papers from the 15th international conference on historical linguistics* (p. 205-219). Philadelphia, PA:John Benjamins Publishing Co.

Janda, R. (2008). Phonologization as the start of dephoneticization or, on sound change and its aftermath: Of extension, generalization, lexicalization, and morphologization. In B. Joseph & R. Janda (Eds.), *The handbook of historical linguistics* (pp. 401-422). Oxford, UK: Blackwell Publishing Ltd.

Jarosz, G. (2006). Richness of the base and probabilistic unsupervised learning in Optimality Theory. In R. Wicentowski & G. Kondrak (Eds.), *Proceedings of the eighth meeting of the ACL special interest group in computational phonology and morphology* (pp. 50-59). New York City.

Johnson, K. (1997). Speech perception without speaker normalization: an exemplar model. In K. Johnson & K. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145-165). New York, NY:Academic Press.

Kabak, B., & Idsardi, W. (2003). Syllabically conditioned perceptual epenthesis. *Proceedings of the Berkeley Linguistic Society*, *29*, 233-245.

Kaun, A. (2004). The typology of rounding harmony. In B. Hayes, R. Kirchner, & D. Steriade

(Eds.), *Phonetically based phonology* (pp. 87-116). New York, NY:Cambridge University Press.

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 115-132). New York, NY:Academic Press.

Keating, P., & Lahiri, A. (1993). Fronted velars, palatalized velars, and palatals. *Phonetica*, *50*, 73-101.

Kenstowicz, M. (1994). *Phonology in generative grammar*. Malden, MA:Blackwell Publishing.

Kingston, J., & de Lacy, P. (2006). *Synchronic explanation*. ms.

Kinkade, M. (1973). The alveopalatal shift in Cowlitz Salish. *International Journal of American Linguistics*, *39*(4), 224-231.

Kiparsky, P. (1995). The phonological basis of sound change. In J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 640–670). Cambridge, MA:Basil Blackwell.

Kiparsky, P. (2006). The amphichronic program vs. evolutionary phonology. *Theoretical Linguistics*, 32, 217-236.

Kiparsky, P. (2008). Universals constrain change; change results in typological generalizations. In J. Good (Ed.), *Linguistic universals and language change* (pp. 23-53). Oxford University Press.

Kirby, S. (1997). Competing motivations and emergence: explaining implicational hierarchies. *Language Typology*, *1*(1), 5-32.

Kochetov, A. (2002). *Production, perception and emergent phonotactic patterns: a case of contrastive palatalization* (L. Horn, Ed.). New York, NY:Routledge.

Krause, S. (1980). *Topics in Chukchee phonology and morphology*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Urbana, IL.

Labov, W. (1972). The recent history of some dialect markers on the island of Martha's Vineyard, Mass. In L. M. Davis (Ed.), *Studies in linguistics in honor of Raven I. Mcdavid*

Jr. U. of Alabama Press.

Labov, W. (1981). Resolving the neogrammarian controversy. *Language*, *57*, 267-309.

de Lacy, P. (2006). *Markedness: Reduction and preservation in phonology*. New York, NY: Cambridge University Press.

Lindblom, B. (1983). Economy of speech gestures. In P. MacNeilage (Ed.), *The production of speech* (pp. 217-245). New York, NY:Springer.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon. Research on speech perception. Technical report no. 6*. Unpublished doctoral dissertation, Indiana University. Bloomington, IN.

Luce, R., & Galanter, E. (1963). Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 191-243). Oxford, UK:Wiley.

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition*, *7*, 485-495.

Maddieson, I., & Precoda, K. (1992). Syllable structure and phonetic models. *Phonology*, *9*, 45-60.

Matthews, W. (1960). *Russian historical grammar*. University of London, Athlone Press.

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

Moreton, E. (1999). Evidence for phonological grammar in speech perception. In *Proceedings of the 14th international congress of phonetic sciences* (pp. 2215-2217). San Francisco.

Neeld, R. (1973). *Remarks on palatalization* (Tech. Rep.). Columbus, OH.

Newman, R., Sawusch, J., & Luce, P. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(3), 873-889.

Niyogi, P., & Berwick, R. C. (1997). A dynamical systems model for language change. *Complex Systems*, *11*, 161-204.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700-708.

Odden, D. (2005). *Introducing phonology*. New York, NY:Cambridge University Press.

Oden, G. C., & Massaro, D. (1978). Integration of featural information in speech perception. *Psychological Review*, *3*, 172-191.

Ohala, J. (1981). The listener as a source of sound change. In C. Masek, R. Hendrick, & M. Miller (Eds.), *Parasession on language and behavior: Chicago linguistics society* (pp. 178-203). Chicago.

Ohala, J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 258-275). New York, NY:Cambridge University Press.

Ohala, J. (1993). The phonetics of sound change. In C. Jones (Ed.), *Historical linguistics: Problems and perspectives* (pp. 237-278). New York, NY:Longman.

Pearl, L. S. (1997). *Necessary bias in natural language learning*. Unpublished doctoral dissertation, University of Maryland. College Park, MD.

Pearl, L. S., & Lidz, J. (2009). When domain-general learning fails and when it succeeds: Identifying the contribution of domain specificity. *Language Learning and Development*, *5*(4), 235-265.

Peterson, G., & Barney, H. L. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America*, *24*(2), 175-184.

Phillips, B. (1984). Word frequency and the actuation of sound change. *Language*, *60*(2), 320-342.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137-158). Philadelphia, PA:John Benjamins Publishing Co.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, T. Rietvelt, & N. Warner

(Eds.), *Laboratory phonology VII* (pp. 101-139). New York, NY:Mouton de Gruyter.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech*, *46*(2-3), 115-154.

Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Attention, Perception, & Psychophysics*, *13*, 253-260.

Pitt, M. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics*, *60*(6), 941-951.

Prince, A., & Smolensky, P. (2004). *Optimality Theory*. Malden, MA:Blackwell Publishing.

Regier, T. (1995). A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics*, *6*(1), 63-88.

Repp, B. (1977). Dichotic competition of speech sounds: the role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 37-50.

Rosch, E. (1977). *Human categorization*. London, UK:Academic Press.

Rosinski, R. R. (1977). Picture-word interference is semantically based. *Child Development*, *48*, 643-647.

Samuel, A. (1982). Phonetic prototypes. *Attention, Perception, & Psychophysics*, *31*, 307-314.

Scobbie, J. M., & Stuart-Smith, J. (2008). Quasi-phonemic contrast and the fuzzy inventory: Examples from Scottish English. In *Contrast: Perception and acquisition: Selected papers from the second international conference on contrast in phonology* (pp. 87-113). Berlin: Mouton de Gruyter.

Shevelov, G. (1979). *A historical phonology of the Ukrainian language*. Heidelberg:Carl Winter Universitatsverlag.

Solé, M.-J. (1992). Phonetic and phonological processes: the case of nasalization. *Language and Speech*, *35*(1,2), 29-43.

Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer et al. (Eds.), *Parallel problem solving from nature VI* (Vol. 1917, pp. 17-26). Berlin:Springer.

Steels, L. (2005). The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. *Connection Science*, *17*(3-4), 213-230.

Stemberger, J. (1990). Wordshape errors in language production. *Cognition*, *35*(2), 123-157.

Tenenbaum, J., & Griffiths, T. (2001). Generalization, similarity and Bayesian inference. *Behavioral and Brain Sciences*, *24*, 629-641.

Tesar, B., & Smolensky, P. (1998). Learnability in Optimality Theory. *Linguistic Inquiry*, *29*(2), 229-268.

Townsend, C., & Janda, L. (1996). *Common and comparative slavic: Phonology and inflection*. Columbus, OH:Slavica Publishers, Inc.

Vapnik, V. (1995). *The nature of statistical learning theory*. New York, NY:Springer-Verlag.

Vitevitch, M., Luce, P., Pisoni, D., & Auer, E. (1999). Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language*, *68*(1-2), 306-311.

Walker, R. (2001). Positional markedness in vowel harmony. *Proceedings of HILP* 5, *12*, 212-232.

Wedel, A. (2004). Category competition drives contrast maintenance within an exemplar-based production/perception loop. In J. Goldsmith & R. Wicentowski (Eds.), *Proceedings of the seventh meeting of the ACL special interest group in computational phonology* (pp. 1-10).

Wedel, A. (2006). Exemplar models, evolution and language change. *The Linguistic Review*, *23*, 247-274.

Wedel, A. (2007). Feedback and regularity in the lexicon. *Phonology*, *24*, 147-185.

Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 95-188). Austin:U. of Texas Press.

Wilson, C. (2006). Learning phonology with substantive bias: an experimental and computational study of velar palatalization. *Cognitive Science*, *30*, 945-982.

Winitz, H., Scheib, M., & Reeds, J. (1972). Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America*, *51*(4(2)), 1309-1317.

Yaeger, M., Labov, W., & Steiner, R. (1972). *A quantitative study of sound change in progress* (Tech. Rep.). Philadelphia: U. S. Regional Survey.

Zilynskyj, I. (1979). *A phonetic description of the Ukrainian language*. Harvard University Press.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Oxford, UK:Addison-Wesley.

APPENDIX

Additional Model Parameters

Token Space

Table A1 here

Locations in F1, F2 space (Hz) for the right-hand vowel context of /k/ tokens in the pre-change lexicon are generated by roughly Gaussian perturbations from the above means. No points are allowed to deviate over two standard deviations from the mean (σ_1= 50; σ_2= 200). The phonetic vowel locations do not change over the course of the simulations, only whether the preceding segment is classified as /k/ or /k^j/.

Lexicon Generation It has been observed that a variety of linguistic units show a very non-uniform distribution: the highest frequency items are observed (in, e.g., a text sample) significantly more often than the next most frequent, and the largest number of different types is found at the lowest rate of occurrence. A distribution in which the absolute frequency of occurrence depends on the relative frequency of occurrence is known as a Zipfian distribution (Zipf, 1949). A particular instantiation of a Zipfian distribution (the Standard Harmonic) is characterized by the following formula

$$f \propto \frac{1}{r}$$

which describes a dependency in which the frequency of a type (*f*) is proportional to its rank frequency (*r*). This means that the second most frequent element will occur half as often as the most frequent element; the third most frequent element will occur one third as often, and so on.

Separate Standard Harmonic distributions were generated for the consonants and vowels in the model; each lexicon instantiated a different random assignment of ranks over each distribution. Subsequently, 70 of each type of syllable (CV, CVC) were randomly generated from the consonant and vowel distributions. This inventory of syllables was in turn randomly sampled to produce a lexicon of 6000 words, 1080 of which were one syllable in length; 2760, two syllables, and 2160, three syllables long.

Simulation Pseudo-Code

```
Main Program

        Generate Lexicon

        Sound Change 0,1:
                kɪ# > k^jɪ# with 75% probability
                All final jers delete (for words longer than 1 syllable)

        Update lexicon, grammar

        #Perception Model
        loop for each Event
                loop for each word
                        loop for each kV sequence
                        Sound Change 2:
                                if (Probability Threshold)
                                        kV > k^jV
                                endif
                        end loop
                end loop
                loop for each word
                        loop for each kV sequence
                        Sound Change 2:
                                if (Probability Threshold)
                                        k^jV > kV
                                endif
                        end loop
                end loop

                if (Event = 1)
                        loop for each word
                                loop for each k^jek sequence
                                Sound Change 3:
                                        k^jek > k^jok with 70% probability
                                end loop
                        end loop
                endif

                Update locations of newly changed tokens
                Update statistics

        #Production model
                loop for each word
                        loop for each {k,k^j}V sequence
                        Entrenchment: Update locations of all tokens
                                New location = average of 10 nearest category neighbors +
                                                Gaussian deviation
                        end loop
                end loop
end loop
```

Probability Threshold:

For a given word containing kV or k^jV:

Calculate Phonotactics

If V is {ɪ, e, or i} then

Calculate Similarity

If V = i then w = 1.0

If V = ɪ then w=.8

If V = e then w=.6

endif

If k > k^j then

Probability Threshold = .08*Phonotactics + w*Similarity

If k^j > k then

Probability Threshold = .08*Phonotactics + (1/w)*Similarity

endif

Calculate Phonotactics:

For a given word $x=..y_{i-2}y_{i-1}qy_{i+1}y_{i+2}..$
where $q = k$ or k^j:

Phonotactics(y,q) = Predictability$_{i-1;i}$(y_{i-1}) + Predictability$_{i-2;i}$(y_{i-2}) + Predictability$_{i+1;i}$(y_{i+1}) + Predictability$_{i+2;i}$(y_{i+2})

Predictability:

Use Maddiesen & Precoda (1992) measure of
predictability: how much more likely segment
y is to occur in position v when segment q
occurs in position w

$$Pd_{vw}(y) = P(y_v \mid q_w) - p(y_v)$$

$$= \frac{P(y_v, q_w)}{P(q_w)} - p(y_v)$$

Estimate the probabilities by the counts from
the lexicon

$$E(Pd_{vw}(y)) = \frac{count(y_v q_w)}{count(q_w)} - \frac{count(y_v)}{n}$$

Calculate Similarity

Likelihood of k^j > k change is 25% that of
k > k^j change

A=.25

S_q = number of tokens in q category
d_{xq} = Euclidean distance between token x and the Center of q in F1, F2 space

N = Normalization factor such that points exactly halfway between the two categories have a 50% probability for either categorization

k^j has been written as C for readability in the
equations

If k > C then

$$P_S(k > C)_x = e^{-N(\frac{d_{xC}}{d_{xk}})\sqrt{\frac{S_C}{S_k}}}$$

Elseif C > k then

$$P_S(C > k)_x = Ae^{-N(\frac{d_{xk}}{d_{xC}})\sqrt{\frac{S_k}{S_C}}}$$

endif

Similarity = P_S

Center

#For a given exemplar cloud of category q

#for natural contexts

$F(q_i)$ = F of following V_i

#for unnatural contexts

$F(q_i)$= F of previous category center

C_1 = average(F1 of tokens of category q)

C_2 = average(F2 of tokens of category q)

Center = (C_1,C_2)

Independent Phoneme changes

Synchronic vowel harmony can be described, under some theoretical frameworks, as a process by which some or all of the features of a trigger vowel spread to one or more target vowels in a given domain, typically with no phonological impact on the consonants of that domain (e.g., Hyman, 2002; Kaun, 2004; Walker, 2001). In this way, mid vowels in proximity to high vowels may become high vowels, unround vowels in proximity to round vowels may become round, etc. The example in (A1) illustrates one such alternation in Chukchee (Krause, 1980). Here an /a/ anywhere in the word induces lowering of /i/ (to [e]), /u/ (to [o]) and /e/ (to [a]); /ə/ does not participate.

(A1) Chukchee (Krause, 1980)

/jara/+/nu/→[jarano]	"as a tent"
/ɣe+rərka+te/→[ɣarərkata]	"with a knife"
/ɣe+titi+ma/→[ɣatetema]	"needle" Comitative

To determine if some kind of independent vowel harmony phoneme change could produce a reversed implication language, let us reformulate the examples of (8) and (9).

First, for simplicity, take a language with only two vowels (o and i), and apply regressive height harmony after the palatalizing phoneme change that converts ki to k^ji. For the hypothetical word in (A2), a vowel harmony change raising o to i (due to the presence of a word-final /i/) acts to increase the number of ki bigrams, reducing the association of /k^j/ with i (see above). Alone, this is not enough to produce an implication reversal, but coupled with the complementary change in (A3) may potentially do so. In (A3), a word-final o lowers i to o, increasing the number of k^jo bigrams, while at the same time decreasing the number of k^ji bigrams.

(A2) koti > kiti

(A3) k^jito > k^joto

The words of (A2) and (A3), in isolation, represent the exact type of vowel swap described above as necessary to produce a reversed implication.

Considering a larger set of word forms – a lexicon –, however, reveals the additional conditions that are required. It would be necessary for *all* words of this hypothetical language to have the exact structure of either (A2) or (A3). Otherwise the reversal pattern will be diluted. Words in which the vowels are originally the same will preserve the natural palatalizing contexts; and words in which the vowels differ, but the consonants are the same (either both /k/, or both /k^j/) will act to flatten the distribution toward uniformity. The exact structure of words in the lexicon and their relative proportions becomes critical to the outcome.

Furthermore, this is a language with only a two vowel inventory. Whether such an impoverished system could be used to represent a full scale – implication reversed or not – raises a new question for the theory. That is, does a learner in this scenario consider the two vowels as representing points on a scale of vowel qualities, prompting them to posit a grammar that effectively places never-encountered vowels in a reversed implication? In the absence of innate, substantive linguistic knowledge, such a scale of vowel qualities would have to be learned from the input. The more vowels and vowel contexts the learner encounters, the more chance they will have to learn such a scale. Once a larger inventory of vowels is added to the scenario depicted in (A2) and (A3), however, the exact conditions required to produce a robustly reversed lexicon multiply – conditions on word structure, lexical distribution, and the vowel harmony process itself which, taken together, significantly reduce the likelihood of such an outcome[11].

Another candidate change that could conceivable produce the reversal outcome is a type of cyclic vowel movement known as a chain shift. It is theoretically possible to construct a series of chain shifts that would essentially swap vowels in the necessary way. For a four vowel system (i,ɪ,e,o), three counter-clockwise shifts in the canonical representations of F1-F2 space would be required to have each vowel of the palatalization hierarchy swap places with its complement. This idealizes the locations of the vowel in symmetric positions, such that each shift moves a vowel to the immediately phonetically adjacent vowel in a counter-clockwise direction. See Figure A1.

Figure A1 approx here

For this to produce the desired effect it must further be the case that the chain shift does not affect the velar segments within the words whose vowels are changing. It would also be necessary for the chain shifts to occur prior to the advent of palatalized variants before non-palatalizing vowels. See (A4) for the sequences that would result from a truly context-free series

---

[11] Bhat (1978) suggests another possibility for effecting an implication flip. This involves a type of phoneme change similar to the third type of change in Table 1, but based entirely on local feature mis-parsing. The vocalic source of palatalization is "absorbed" into the palatalized consonant, its frontness features reattributed to the frontness on the consonant in a sequence such as that in (1).

(1) ki > k^ji > k^jo

The upshot of this type of change is a lexicon in which palatal variants only occur before non-front vowels. However, for this to be categorically (or quasi-categorically) true, the second part of the change in (1) must be uniform. Assuming that the first change in the sequence is also uniform, removes front vowel environments entirely from the distribution. Thus, both palatal and plain variants occur before o, and *neither* variant occurs before i: a contrast outcome. In the non-uniform change version, both ki sequences and k^ji sequences persist, resulting once again in a system of contrast. Although Bhat reports several languages in which palatal variants only appear before non-front vowels, it is not clear from the descriptions whether these actually represent implication reversals. Such data (if sufficiently detailed) could serve as a good independent test case for the model, providing an assessment of whether or not the above analysis captures the relevant dimensions of variation.

of chain shifts. The net result is that palatal variants never occur before i, and that the two phonemes are in contrast before all other vowels (effectively, neutralization to /k/ before i). If palatal variants have already spread to o contexts, however, the chain shifts will include the context in (A5), adding back in the i context for palatals, and producing no net effect on the system.

(A4) k^{j}i>k^{j}o   k^{j}ɪ>k^{j}e   k^{j}e>k^{j}ɪ
     ki>ko   kɪ>ke   ke>kɪ   ko>ki

(A5) k^{j}o>k^{j}i

This scenario can be argued to be fairly implausible as well. Although completed chain shifts are so-called based on their regularity, chain shifts in progress can be shown to depend on the fine grained phonetic details of particular words (e.g., Yeager et al., 1972). Thus a particular degree of raising or fronting on a vowel undergoing the shift will depend on whether the neighboring segments in a particular word are antagonistic to the direction of the change or not. By the same token, if a vowel shift eventually proceeds despite phonetic incompatibility, it might be expected that the neighboring segments will show some effect of the change as well.

Our requirements on the exact shape of the chain shift may also prove to be implausible. Descriptions of the Northern Cities chain shift (Yeager et al., 1972), involve the fronting and raising of a low central vowel to instigate the chain, leaving a gap into which the o vowel can move, in turn leading to shifts in the ʊ and u positions. Meanwhile, i, under pressure from the raising a, can back in the direction of the former u position. This is more or less consistent with the schematic in Fig. A1. ɪ and e, however, are described as falling and backing in response to the pressure from the raising a vowel, whereas the trajectory above requires them to raise to i and ɪ, respectively. Especially in an inventory missing a mid central vowel, mid vowels might be expected to move to this central location rather than follow a strictly counter-clockwise chain. The additional requirement that a similar series of shifts occur multiple times reduces the probability of the entire trajectory yet further.

Table 1: Full set of phoneme changes. Columns 2 and 3 contain example words to illustrate the changes. Bold font indicates a change in underlying form from the previous stage.

Stage 0	/dakɪ/→[dak^{j}ɪ]	/taki/→[tak^{j}i]	Phonetic Palatalization Articulatory Ease
	/dakʊ/→[dakʊ]	/kem/→[k^{j}em]	
	/kek/→[k^{j}ek]	/strakɪlit/→[strak^{j}ɪlit]	
		/kisum/→[k^{j}isum]	
		/zoket/→[zok^{j}et]	
		/kɪb/→[k^{j}ɪb]	
	Loss of Final Jer		
Stage 1	**/dak^{j}/→[dak^{j}]**	/taki/→[tak^{j}i]	New Phoneme Category Initiated (minimal pairs)
	/dak/→[dak]	/kem/→[k^{j}em]	
	/kek/→[k^{j}ek]	/strakɪlit/→[strak^{j}ɪlit]	
		/kisum/→[k^{j}isum]	
		/zoket/→[zok^{j}et]	
		/kɪb/→[k^{j}ɪb]	
Stage 2	/dak^{j}/→[dak^{j}]	**/tak^{j}i/→[tak^{j}i]**	Re-categorization Re-analysis (bi-directional)
	/dak/→[dak]	**/k^{j}em/→[k^{j}em]**	
	/kek/→[k^{j}ek]	**/strak^{j}ɪlit/→[strak^{j}ɪlit]**	
		/k^{j}isum/→[k^{j}isum]	
		/zok^{j}et/→[zok^{j}et]	
		/k^{j}ɪb/→[k^{j}ɪb]	
		/kisum/→[k^{j}isum]	
		/kɪb/→[k^{j}ɪb]	
Stage 3	/dak^{j}/→[dak^{j}]	/tak^{j}i/→[tak^{j}i]	Contrast Maintenance + Re-analysis
	/dak/→[dak]	/k^{j}em/→[k^{j}em]	
	/k^{j}ok/→[k^{j}ok]	/strak^{j}ɪlit/→[strak^{j}ɪlit]	
		/kisum/→[k^{j}isum]	
		/zok^{j}et/→[zok^{j}et]	
		/kɪb/→[k^{j}ɪb]	

Table 2: Possible Phoneme change Trajectory over Lexicon Fragment.
Changes are applied from left to right in time. Bold text indicates a change in the underlying form from the previous stage. Gray cells indicate that the context for the given change was present, but that the change did not occur (by chance).

Stage 0	Stage 1	Stage 2	Stage 2	Stage 3
dʊgkɪ	**dʊgk^j**	dʊgk^j	dʊgk^j	dʊgk^j
kekbi	kekbi	**k^{j}ekbi**	**kekbi**	kekbi
kɪdma	kɪdma	**k^{j}ɪdma**	k^{j}ɪdma	k^{j}ɪdma
zɪdbɪkɪd	zɪdbɪkɪd	zɪdbɪkɪd	zɪdbɪkɪd	zɪdbɪkɪd
nɪkɪ	**nɪk**	nɪk	nɪk	nɪk
mamuki	mamuki	**mamuk^{j}i**	mamuk^{j}i	mamuk^{j}i
kiztidug	kiztidug	**k^{j}iztidug**	k^{j}iztidug	k^{j}iztidug
pikɪd	pikɪd	**pik^{j}ɪd**	pik^{j}ɪd	pik^{j}ɪd
mikib	mikib	**mik^{j}ib**	mik^{j}ib	mik^{j}ib
dikek	dikek	**dik^{j}ek**	dik^{j}ek	**dik^{j}ok**
mukɪ	**muk^j**	muk^j	muk^j	muk^j
zobɪkim	zobɪkim	zobɪkim	zobɪkim	zobɪkim
sʊkɪs	sʊkɪs	**sʊk^{j}ɪs**	**sʊkɪs**	sʊkɪs
kɪngan	kɪngan	kɪngan	kɪngan	kɪngan
keknʊn	keknʊn	**k^{j}eknʊn**	k^{j}eknʊn	**k^{j}oknʊn**
mɪkdodkek	mɪkdodkek	mɪkdodkek	mɪkdodkek	mɪkdodkek
dʊkguki	dʊkguki	**dʊkguk^{j}i**	**dʊkguki**	dʊkguki
tobugkɪ	**tobugk^j**	tobugk^j	tobugk^j	tobugk^j
kɪbopkɪ	**kɪbopk**	**k^{j}ɪbopk**	k^{j}ɪbopk	k^{j}ɪbopk

Table 3: Phonotactics over Lexicon Fragment Subsequent to Phoneme changes.
Columns 2, 4, 6, 8 represent the four environments for which phonotactic relationships are computed by the learner. The same fragment appears in columns 1, 3, 5, and 7, but is re-ordered for the purpose of grouping partially identical phonological environments.

	..kʲ__..		..kʲ__ __..		..__kʲ..		..__ __kʲ..
kʲɪpsɪs	kʲɪ	kʲɪpsɪs	..kʲ__p..	kʲɪpsɪs	..#kʲ..	kʲɪpsɪs	∅
sʊmkʲɪp		sʊmkʲɪp		kʲɪ		kʲɪ	
toskʲɪp		toskʲɪp		kʲɪsɪs		kʲɪsɪs	
kʲɪ		kʲɪpset		kʲɪpkon		kʲɪpkon	
kʲɪsɪs		tomkʲɪpbe		kʲɪpset		kʲɪpset	
kʲɪpkon		kʲepʊssu		kʲɪnem		kʲɪnem	
kʲɪpset		sekʲep		kʲɪsʊp		kʲɪsʊp	
kʲɪnem		kʲʊpʊs		kʲepʊssu		kʲepʊssu	
kʲɪsʊp		kʲʊpʊssʊt		kʲʊpʊs		kʲʊpʊs	
tomkʲɪpbe		kʲɪpkon		kʲʊpʊssʊt		kʲʊpʊssʊt	
kʲedame	kʲe	kʲɪ	..kʲ__#..	kʲedame		kʲedame	
putɪpkʲe		putɪpkʲe		kʲʊsu		kʲʊsu	
kʲepʊssu		kʲɪsɪs	..kʲ__s..	putɪpkʲe	..pkʲ..	sʊmkʲɪp	..ʊ__kʲ..
sekʲep		kʲɪsʊp		tomkʲɪpbe	..mkʲ..	toskʲɪp	..o__kʲ..
kʲʊpʊs	kʲʊ	kʲʊsu		sʊmkʲɪp		tomkʲɪpbe	
kʲʊpʊssʊt		kʲɪnem	..kʲ__n..	toskʲɪp	..skʲ..	putɪpkʲe	..ɪ__kʲ..
kʲʊsu		kʲedame	..kʲ__d..	sekʲep	..ekʲ..	sekʲep	..s__kʲ..

Table 4: Model Output Grammar Types
For simplicity, the current model is assumed to produce more or less implication compliant types (cell 1); therefore no additional phoneme changes are needed in column 1.

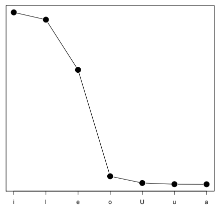
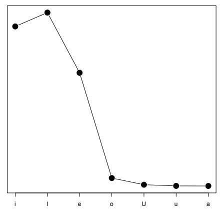
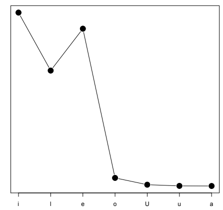
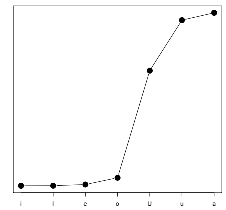
	Implication Compliant	Partial Violation		Complete Implication Violation
Implicational Hierarchy				
Necessary Change		Loss of k^ji Gain of k^jɪ	Gain of kɪ	Swap Contexts of A
Possible Phoneme change		k^ji>k^jɪ	Large Borrowing	Context-free Vowel Quality Reversal a>i i>a u>ɪ ɪ>u ʊ>e e>ʊ

Table A1: Adult male speaker, mean formant frequencies in Hertz.

	i	ɪ	e	u	ʊ	o	a
F1	300	470	580	330	500	600	840
F2	2440	2000	1930	1010	1260	980	1220

Figure 1
The Model

Figure 2
Category tokens in F1-F2 phonetic space (Hz) given by following vowel.
Black circles are /k/ tokens. Single large white circle represents initiation of /k^j/ category at start of simulations.

Figure 3
Lexicon A
Evolution of 2 categories over multiple phoneme change 'Events' (from top left to bottom right): 0, 9, 19, 50. x-axis: F2; y-axis: F1. White circles represent k^j tokens; black circles, k tokens.

Figure 4
Lexicon A
Pre-vocalic Distribution of 2 categories over multiple phoneme change 'Events' (from top left to bottom right): 0, 9, 19, 50; white: k^jV bigram, black: kV bigrams. x-axis arranged by implication order.

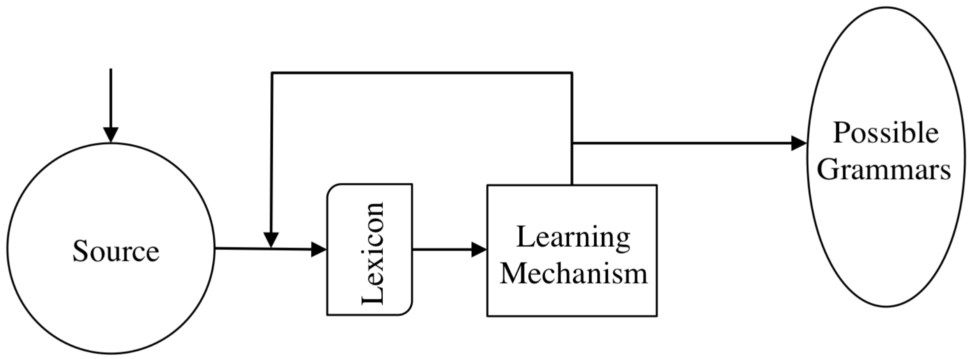
Figure 5
The three lexicons compared
Top row: phonetic distribution at the beginning of the simulations (Event 0); bottom row: phonological distribution by Event 50.

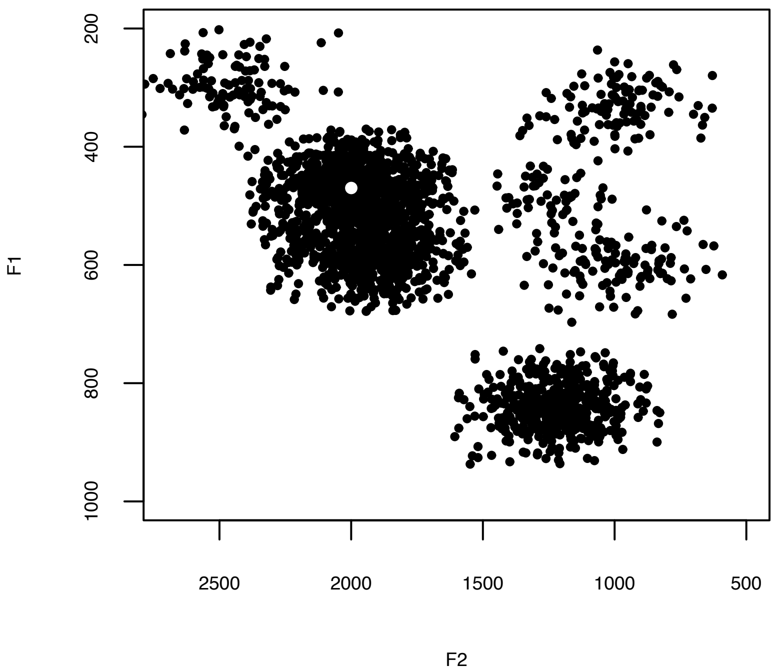
Figure 6
Number of occurrences (out of 1000) in which threshold predictability is reached for each of four phonotactic associations (four panels), over all segments (0=word end; 1=word beginning; C=k^j).
Threshold predictability is calculated locally as 60% of the strongest phonotactic in the given language. Note that the y-axis of each of the panels is scaled to the maximum value for that panel. Each simulation runs for 6 Events.

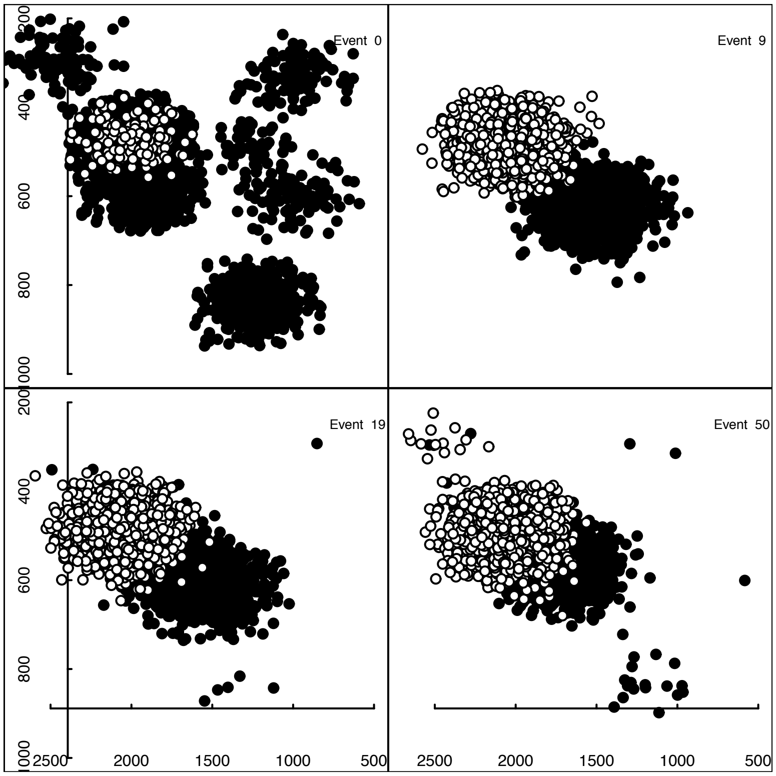
Fig. 7
Average Proportion of occurrence with k^j versus k, by vowel, in implicational order

Fig. 8
Difference in Proportion (Degree of Violation) between neighboring vowels (when positive).
Number of instances of gradient implicational violations appears next to boxplot for each pair.

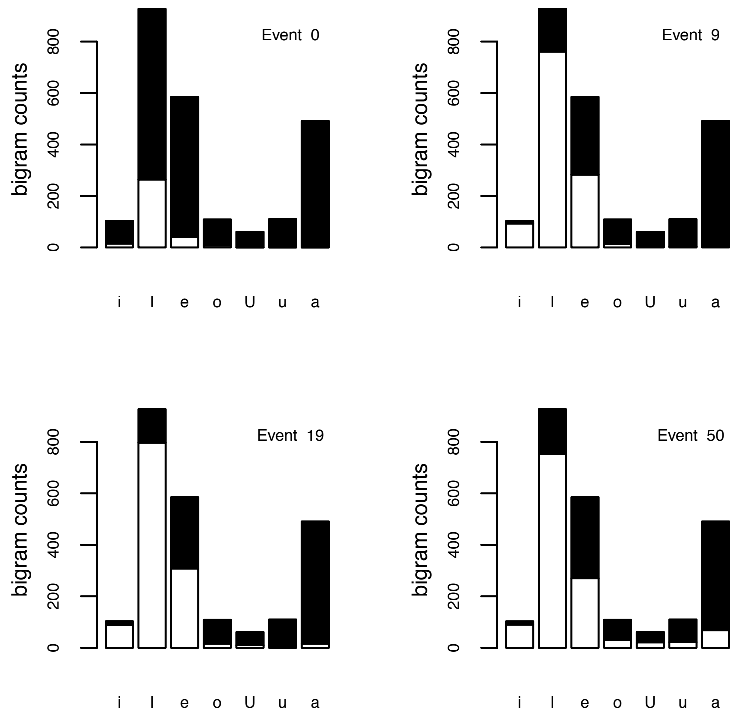
Fig. A1
Hypothetical chain shift for implicational reversal

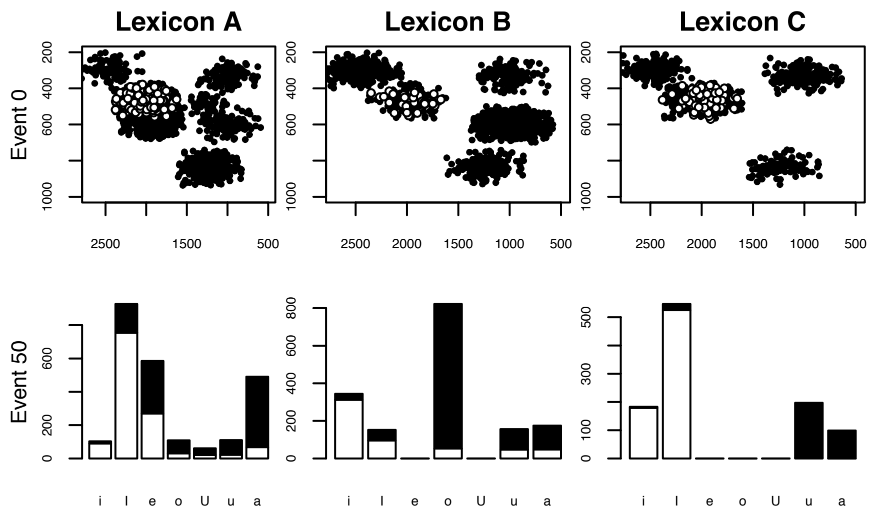

Source
Lexicon
Learning
Mechanism
Possible
Grammars

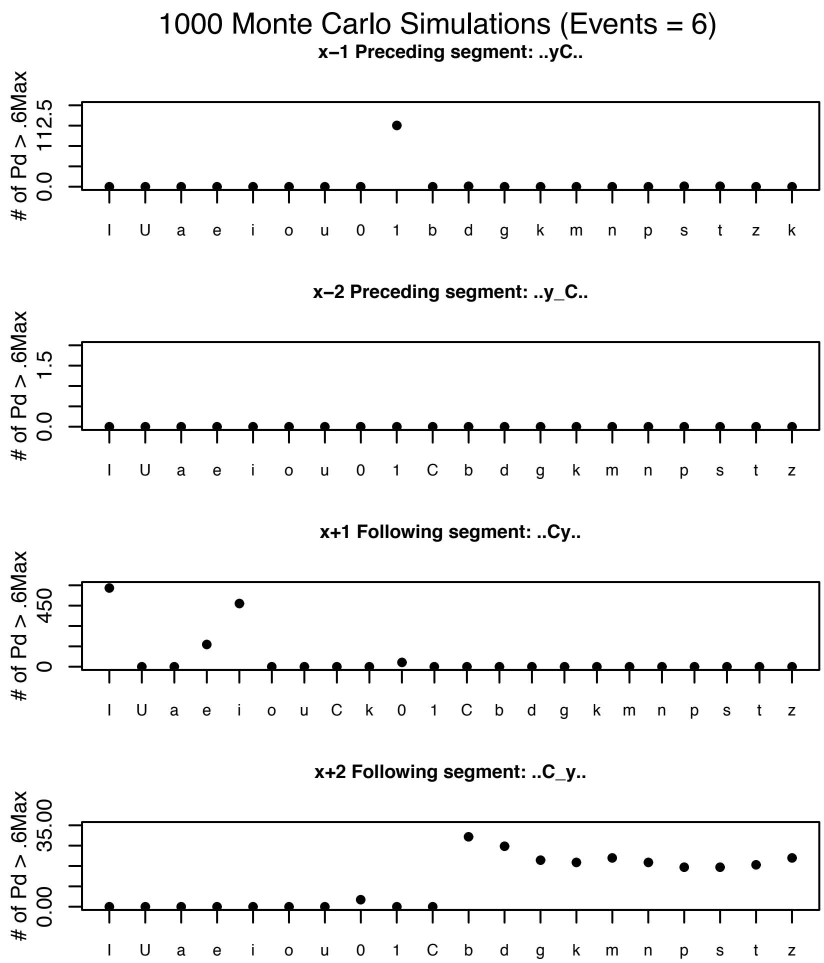

200
400
600
800
1000
F1
2500
2000
1500
1000
500
F2

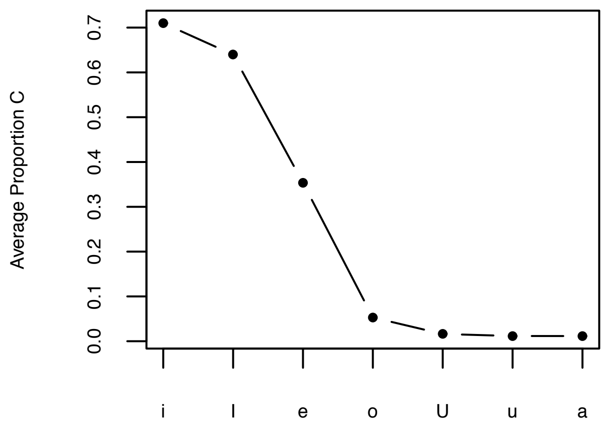


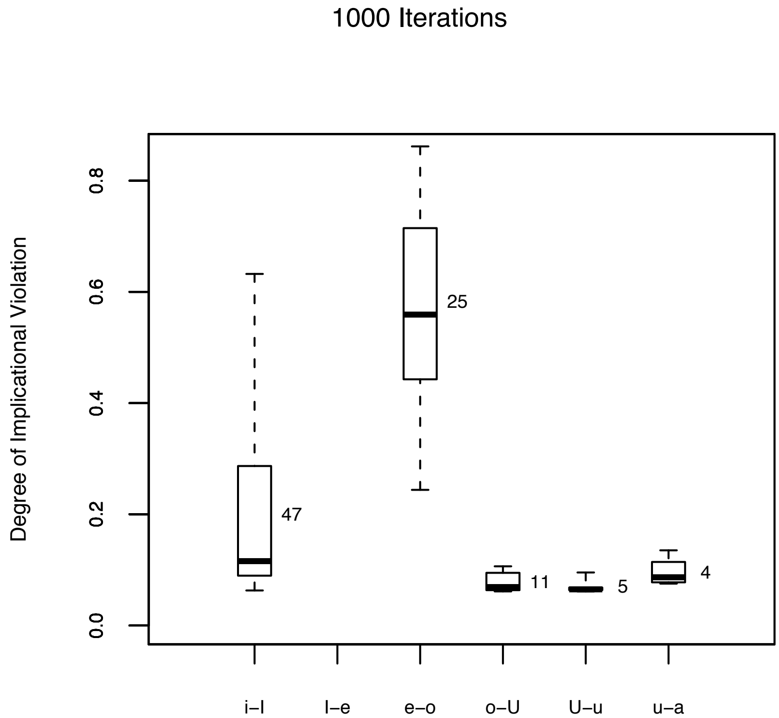
Event 0
Event 9
Event 19
Event 50
200
400
600
800
1000
2500
2000
1500
1000
500

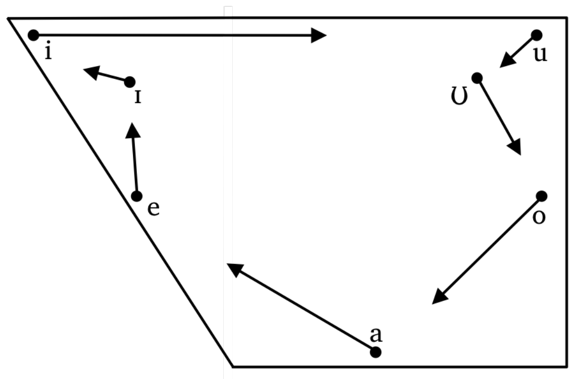

Event 0
Event 9
Event 19
Event 50
bigram counts
0
200
400
600
800
i I e o U u a


Lexicon A
Lexicon B
Lexicon C
Event 0
Event 50
2500
1500
500
2000
1000
i
I
e
o
U
u
a


1000 Monte Carlo Simulations (Events = 6)
x−1 Preceding segment: ..yC..
# of Pd > .6Max
0.0
112.5
I U a e i o u 0 1 b d g k m n p s t z k
x−2 Preceding segment: ..y_C..
# of Pd > .6Max
0.0
1.5
I U a e i o u 0 1 C b d g k m n p s t z
x+1 Following segment: ..Cy..
# of Pd > .6Max
0
450
I U a e i o u C k 0 1 C b d g k m n p s t z
x+2 Following segment: ..C_y..
# of Pd > .6Max
0.00
35.00
I U a e i o u 0 1 C b d g k m n p s t z


Average Proportion C
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
i
I
e
o
U
u
a


1000 Iterations
Degree of Implicational Violation
0.0
0.2
0.4
0.6
0.8
47
25
11
5
4
i–I
I–e
e–o
o–U
U–u
u–a


i
ɪ
e
a
o
ʊ
u