Variation and change in English noun/verb pair stress: Data and dynamical systems models

Morgan Sonderegger*
morgan@cs.uchicago.edu

Partha Niyogi†
niyogi@cs.uchicago.edu

1 Introduction

In every language, change is ubiquitous and variation is widespread. Their interaction is key to understanding language change because of a simple observation: every linguistic change seems to begin with variation, but not all variation leads to change. What determines whether, in a given linguistic population, a pattern of variation leads to change or not? This is essentially the *actuation problem* (Weinreich et al., 1968),[1] which we rephrase as follows: why does language change occur at all, why does it arise from variation, and what determines whether a pattern of variation is stable or unstable (leads to change)? This paper addresses these questions by combining two approaches to studying the general problem of why language change occurs: first, building and making observations from datasets, in the tradition of sociolinguists and historical linguists (such as Labov and Wang); second, building mathematical models of linguistic populations, to model the diachronic, population-level consequences of assumptions about the process of language learning by individuals (Niyogi and Berwick, 1995, *et seq.*; Niyogi, 2006).

*Department of Computer Science University of Chicago.
†Departments of Computer Science and Statistics, University of Chicago.
**We thank two anonymous reviewers for comments on an earlier draft of this chapter, John Goldsmith, Jason Riggle, and Alan Yu for insightful discussion, and Max Bane for both. Audiences at LabPhon 11, the University of Chicago, and Northwestern University provided useful feedback.

[1] "Why do changes in a structural feature take place in a particular language at a given time, but not in other languages with the same feature, or in the same language at other times?" (p. 102)

Pattern	N	V	Examples
$\{1, 1\}$	$\acute{\sigma}\sigma$	$\acute{\sigma}\sigma$	*anchor*, *fracture*, *outlaw*
$\{1, 2\}$	$\acute{\sigma}\sigma$	$\sigma\acute{\sigma}$	*consort*, *protest*, *refuse*
$\{2, 2\}$	$\sigma\acute{\sigma}$	$\sigma\acute{\sigma}$	*cement*, *police*, *review*

Table 1: English noun/verb pair stress patterns.

We describe the diachronic dynamics of an English stress shift, based on a diachronic dataset (1600–2000) which shows both variation and change. This stress shift has several interesting properties which can be explored using computational models. We focus here on a pattern characterizing much language change, the existence of periods of long-term stability punctuated by periods of change. We discuss several proposed causes of change from the literature (listener-based misperception, word frequency, analogy) and their application to our dataset; we then link observed dynamics and proposed causes by determining the diachronic dynamics of three models of learning by individuals in a linguistic population.[2] Based on these models, we argue that bifurcations in the dynamics of linguistic populations are a possible explanation for actuation, and that the presence or absence of bifurcations can be used to evaluate proposed mechanisms of language change.

2 Data

The data considered here are English disyllabic noun-verb pairs such as *convict*, *concrete*, *exile*, referred to as *N/V pairs* throughout. As a rough count of the number of N/V pairs in current use, 1143 are listed in CELEX (Baayen et al., 1996).[3] N/V pairs are a productive class (*YouTube*, *google*).

All current N/V pairs for which N and V have categorical stress follow one of the 3 patterns shown in Table 1.[4] The fourth logically possible pattern, {2,1}, does not occur; as discussed below, this patttern is also never observed diachronically. At any given time, variation exists in the pronunciation of some N/V pairs, e.g. *research*, *address* in present-day American English.

[2] These models are sampled from a larger project (Sonderegger, 2009; Sonderegger and Niyogi, 2010), whose goal is to determine which model properties lead to dynamics consistent with the stress data, and with observations about variation and change more generally.

[3] The number of N/V pairs in current use depends on the method used to count. Many examples are clear, but others have rarely-used N or V forms (e.g. *collect*) which are still listed in dictionaries.

[4] We use curly brackets to denote N and V stress, where 1=initial stress and 2=final stress.

Variation and change in the stress of N/V pairs have a long history. Change in N/V pair stress was first studied in detail by Sherman (1975), and subsequently by Phillips (1984). Sherman (1975) found that many words have shifted stress since the first dictionary listing stress appeared (1570), largely to $\{1, 2\}$.[5] On the hypothesis that this was lexical diffusion to $\{1, 2\}$, he counted 149 pairs listed with $\{1, 2\}$ or possible $\{1, 2\}$ pronunciation in two contemporary dictionaries, one British and one American, and examined when the shift for each N/V pair took place. We call these 149 words List 1 (Appendix A). Sherman found the stress of all words in List 1 for all dictionaries listing stress information published before 1800, and concluded that many words were $\{1, 2\}$ by 1800, and those that were not must have shifted at some point by 1975. We will revisit the hypothesis of lexical diffusion to {1,2} below, after examining the dynamics of an expanded dataset.

Stability of {1,1}, {2,2}, & {1,2} Because Sherman's study only considers N/V pairs which are *known* to have changed to {1,2} by 1975, it does not tell us about the stability of the {1,1}, {2,2}, and {1,2} pronunciations in general. Over a random set of N/V pairs in use over a fixed time period, is it the case that most pairs pronounced {1,1} and {2,2} shift stress to {1,2}?

List 2 (Appendix B) is a set of 110 N/V pairs, chosen at random from all N/V pairs which (a) have both N and V frequency of at least 1 per million in the British National Corpus; (b) have both N and V forms listed in a dictionary from 1700 (Boyer, 1700); (c) have both N and V forms listed in a dictionary from 1847 (James and Molé, 1847). These criteria serve as a rough check that the N and V forms of each word have been in use since 1700.

In List 2, Only 11.8% of the words have changed stress at all from 1700–2007. Those stress shifts observed are mostly as described by Sherman, from $\{2, 2\}$ to $\{1, 2\}$, and mostly for words from List 1. But this quick look suggests that when the set of *all* N/V pairs is sampled from over a 300 year period, most words do not change stress: $\{1, 1\}$, $\{1, 2\}$, and $\{2, 2\}$ are all "stable states," to a first approximation. From this perspective, both sides of the actuation problem are equally puzzling for the dataset: why do the large majority of N/V pairs not change, and what causes change in those that do?

2.1 Diachronic: Dictionary data

To get a better idea of the diachronic dynamics, Sherman's data on N/V stress for List 1 words from 33 British dictionaries were extended to the present using 29 additional British and 14 additional

[5] However, most words are not first listed until 1700 or later.

American dictionaries, published 1800–2003.[6] Words from List 1 were used rather than a list of N/V pairs controlled for first attestation and non-zero frequency (such as List 2) for two reasons. First, we wish to use the large dataset already collected by Sherman for List 1 pronunciations up to 1800. Second, we are interested in the dynamics of *change*, and would therefore like to focus on words which have changed by the present. Because, most pairs do not change stress over time and most change is to {1,2}, List 1 will include most pairs which have undergone a stress shift.

For the 149 N/V pairs of List 1 in 76 dictionaries, each of N and V was recorded as 1 (initial stress), 2 (final stress) 1/2 (both listed, 1 first) 2/1 (both listed, 2 first), 1.5 (level stress), or 0 (not listed). We assume 1/2, 1.5, and 2/1 reflect variation in the population, either due to variation within individuals (e.g. the dictionary's author(s)) or variation across individuals (each using initial or final stress exclusively). At a given time, the N or V forms for many words in List 1 are rare, archaic, or not in use. {2,1} is never observed.

Changes in individual N/V pairs' pronunciations can be visualized by plotting the moving average of their N and V form stresses. To represent averages of reported stresses on a scale, we need to map reported stresses s as numbers $f(s)$ in $[1, 2]$. We use

$$f(1) = 1, \quad f(2) = 2, \quad f(1/2) = f(2/1) = f(1.5) = 1.5$$

This measure overestimates variation between 1 and 2 by interpreting 1/2 and 2/1 as meaning equal variation between 1 and 2.[7]

For a word w at time t, the average of pronunciations reported in the time window $(t-25, t+25)$ (years) was plotted if at least one dictionary in this time window listed pronunciation data for w. So that the trajectories would reflect change in one dialect of English, only data from British dictionaries was used. Figs. 1–2 show a sample of the resulting 149 stress vs. time trajectories.[8]

Four types of complete stress shift, defined as a trajectory moving from one endpoint ($\{1, 1\}$, $\{1, 2\}$, or $\{2, 2\}$) to another, are observed, ordered by decreasing frequency in Table 2. The types differ greatly in frequency: {2,2}→{1,2} is by far the most common, while there are only 1–2 clear examples of {1,2}→{2,2}. For both the {1,1} and {2,2} patterns, change to {1,2} occurs

[6] The dictionary list is in (Sonderegger, 2009); the stress data are available on the first author's web page (currently `people.cs.uchicago.edu/~morgan`).

[7] In fact, dictionary authors often state that the first listed pronunciation is "primary," so that 1/2, 2/1, and 1.5 could represent different types of variation in the population, in view of which we might want to set $f(1/2) < 0.5$ and $f(2/1) > 0.5$. In practice, 1/2 and 2/1 are uncommon enough that trajectories plotted with $f(1/2)$ and $f(2/1)$ changed look similar, at least with respect to the qualitative terms in which we describe trajectory dynamics below.

[8] All trajectories are given in (Sonderegger, 2009), and posted on the first author's web page.

Change	Examples
$\{2,2\} \rightarrow \{1,2\}$	*concert*, *content*, *digest*, *escort*, *exploit*, *increase*, *permit*, *presage*, *protest*, *suspect*
$\{1,1\} \rightarrow \{1,2\}$	*combat*, *dictate*, *extract*, *sojourn*, *transfer*
$\{1,2\} \rightarrow \{1,1\}$	*collect*, *prelude*, *subject*
$\{1,2\} \rightarrow \{2,2\}$	*cement*

Table 2: Observed types of complete stress shift, ordered by decreasing frequency of occurrence.

	V=1	V=var	V=2
N=1	7.1	4.6	57.1
N=var	0	2.2	7.1
N=2	0	0	21.8

Table 3: Distribution (in %) of data with both N and V stresses listed. "Var" means 1.5, 1/2, or 2/1.

more frequently than change from {1,2}. Change directly between $\{1,1\}$ and $\{2,2\}$ never occurs. A sample of each type is shown in Fig. 1.

For a given N or V stress trajectory, variation—a moving average value greater than 1 and less than 2—could either be due to dictionary entries reporting variation, or a mix of dictionary entries without variation reporting (exclusively) initial or final stress. To give an idea of how often variation is reported in individual dictionary entries, Table 3 shows the percentages of entries (with both N and V stresses listed) reporting variation in N, V, or neither. Variation occurs within N or V in 13.9% of entries, but variation in both N and V at once is relatively uncommon (2.2% of entries).

What is the diachronic behavior of the variation observed in the stress trajectories? Examining all trajectories, we can make some impressionistic observations. Short-term variation near end-points (*converse*; Fig. 2(a)), is relatively common. Long-term variation in one of the N or V forms (*exile*; Fig. 2(b)), is less common; long-term variation in both the N and V forms at once (*rampage*; Fig. 2(c)), is rare.

{2,1} is never observed in the dataset, and we argue it is in fact "unstable" in the following sense. Entries "near" {2,1}, such as (N=2/1,V=1/2) are very rare (9 entries), and are scattered across different words and dictionaries. This means that the few times the N form of an N/V pair comes close to having a higher probability of final stress than the V form, its trajectory quickly changes so this is no longer the case. In the language of dynamical systems (§4.1), this suggests the region $\text{pron}_N > \text{pron}_V$ contains an unstable fixed point (one which repels trajectories), {2,1}.

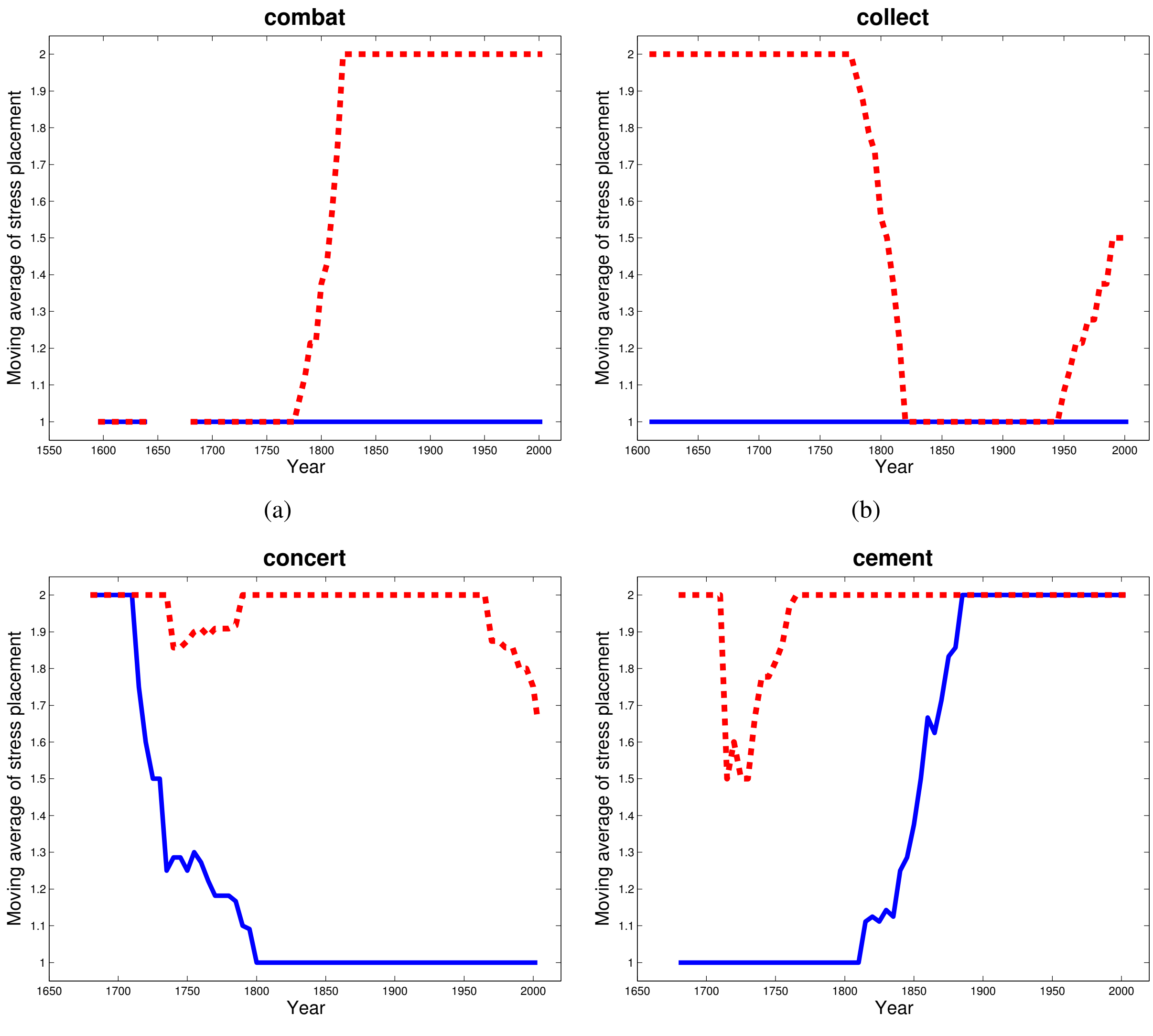
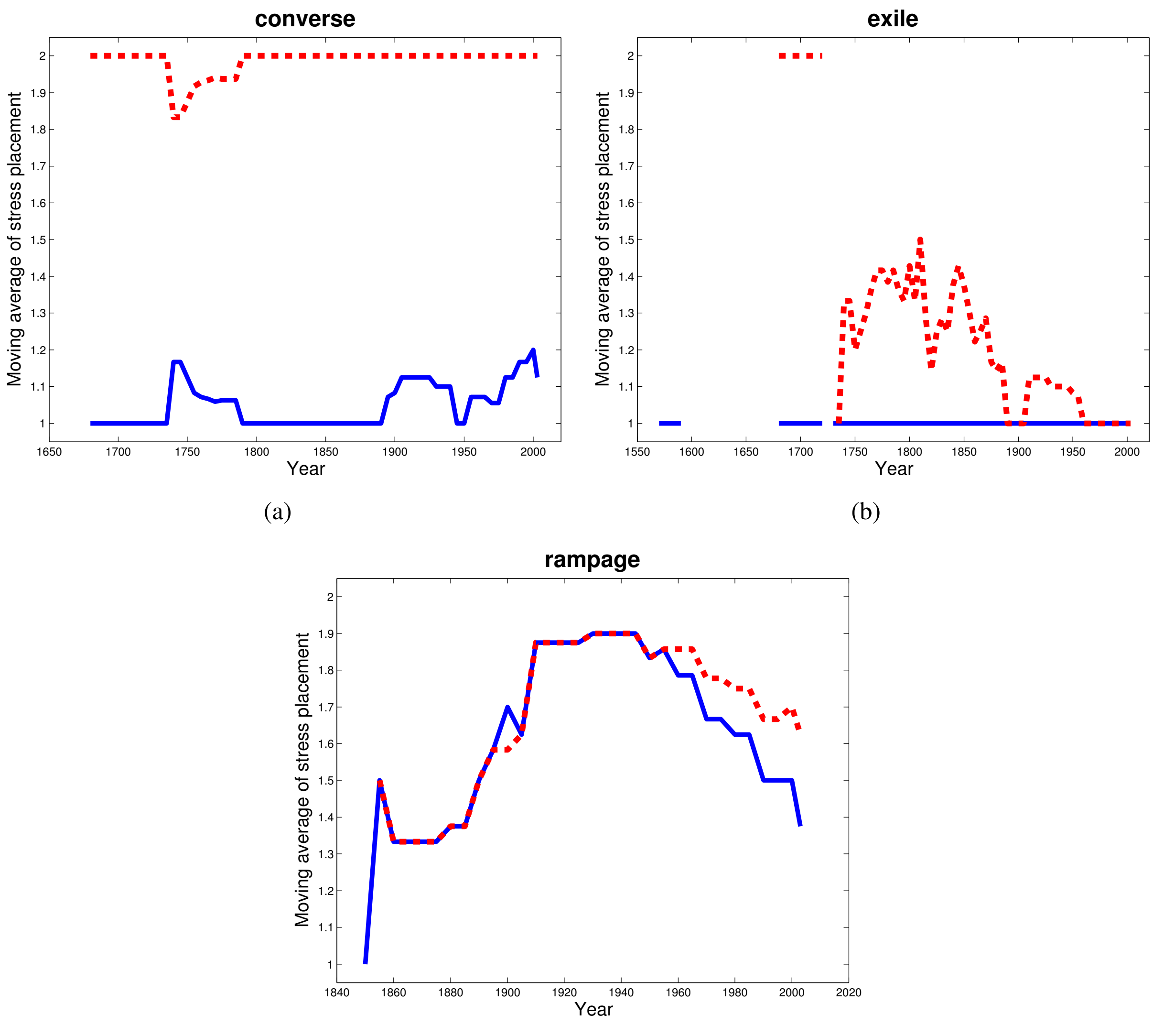


Figure 1: Sample trajectories 1: change between endpoints. Solid/dotted lines are moving averages of N/V stress.



(a) (b)

(c)

Figure 2: Sample trajectories 2. (a) Short-term variation; (b) long-term variation in the V form; (c) long-term variation in both N and V forms. Solid/dotted lines are moving averages of N/V stress.

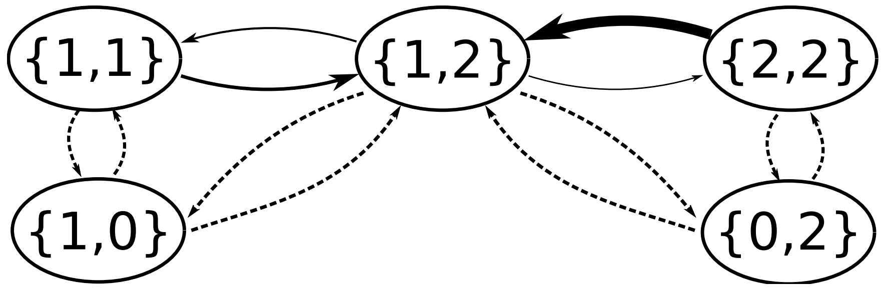


Figure 3: Schematic of observed changes. Each oval represents a stable state: {1,1}, {1,2}, and {2,2} are the observed N/V pair stress patterns, and {1,0} and {0,2} indicate disyllabic words without V and N forms, respectively. Solid lines indicate observed N/V pair stress shifts, with line thickness indicating the relative frequency of each shift; e.g. {2,2}$\rightarrow${1,2} is the most frequent and {1,2}$\rightarrow${2,2} the least frequent. Dotted lines indicate all ways in which an N or V form can come into or fall out of use.

We can summarize the observed diachronic facts as follows:

1. {1,1}, {1,2}, {2,2} are "stable states", but short-term variation around them often occurs.
2. Long-term variation occurs, but rarely in both N and V forms simultaneously.
3. Trajectories largely lie on or near a 1D axis in the 2D $(\text{pron}_N, \text{pron}_V)$ space: $\{1,1\} \leftrightarrow \{1,2\} \leftrightarrow \{2,2\}$. Both variation and change take place along this axis.
4. Changes to {1,2} are much more common than changes from {1,2}.
5. {2,1} never occurs, and is an "unstable state".

Returning to the question of what kind of change is taking place, we see that to a first approximation and restricted to List 1, Sherman was correct: most change takes place to $\{1,2\}$. But taking into account that change from $\{1,2\}$ also occurs, and that most words in stable states never change, the diachronic picture is more completely schematized as in Fig. 3. The observed dynamics are thus more complicated than diffusion to $\{1,2\}$. To understand their origin, we consider below (§ 3) proposed mechanisms driving stress shift in N/V pairs.

2.2 Synchronic: Radio data

We can infer from the dictionary data that significant population-level variation exists in the pronunciation of many N/V pairs at a given time. However, to build realistic models, we must also

Word	# N=1	# Var	# N=2
research	9	6	2
perfume	2	3	4
address	2	2	1

Table 4: Summary of radio pronunciation data (see text).

know whether pronunciation variation exists in individuals or not: do individuals learn gradient (a probability $\alpha \in [0, 1]$ of using one form versus another) or categorical (each speaker uses one form exclusively) forms? We call these options *within-speaker* and *between-speaker* variation.[9]

One place to check the type of variation is on the radio, by observing how an individual speaker pronounces different tokens of words known to show variation at the population level. For a sample of 34 stories from National Public Radio, the American public radio network, Table 4 lists the number of speakers (31 total, 18 male) who pronounced the noun form of *research*, *address*, or *perfume*, exclusively with initial stress, exclusively with final stress, or used both. Each speaker listed for a word used it at least 5 times.[10]

Within-speaker variation thus does occur for N/V pairs, at least in this relatively small dataset. This finding has important consequences for modeling. As has been pointed out in both dynamical systems (Niyogi, 2006) and other computational models of language change (e.g. Liberman, 2000; Troutman et al., 2008), the choice of whether learners' target is a gradient or categorical form profoundly affects the population-level dynamics.

Based on the radio data, we can also make an observation about the structure of within-speaker variation for modeling: although within-speaker variation exists, 2/3 of speakers show no variation at all. This could be taken to suggest that learners are not simply probability matching (assuming their input includes both N=1 and N=2 examples), and that the learning procedure can terminate in gradient *or* categorical output, given gradient input. We do not pursue this possibility in the models presented below.

[9]The terminology is slightly misleading because the structure of variation (the α stored) differs between speakers in "within-speaker" variation as well.

[10]See (Sonderegger, 2009) for details, including the list of stories.

3 Motivations for change

We outline several proposed types of causes of phonological change, and for each one explore its relevance for the observed diachronic dynamics of N/V pair stress.

3.1 Mistransmission

An influential line of research holds that many sound changes are based in asymmetric transmission errors: because of articulatory factors (e.g. coarticulation), perceptual biases (e.g. confusability between sounds), or ambient distortion between production and perception, listeners systematically mishear some sound α as β, but rarely mishear β as α.[11] Such asymmetric mistransmission is argued to be a necessary condition for the change $\alpha \rightarrow \beta$ at the population level, and an explanation for why the change $\alpha \rightarrow \beta$ is common, while the change $\beta \rightarrow \alpha$ is rarely (or never) observed. Mistransmission-based explanations were pioneered by Ohala (1981, et seq.), and have been the subject of much recent work (reviewed by Hansson, 2008)

Although N/V pair stress shifts are not sound changes, their dynamics are potentially amenable to mistransmission-based explanation. There is significant experimental evidence for perception and production biases in English listeners consistent with the most commonly-observed diachronic shifts ({2,2}, {1,1}→{1,2}). English listeners strongly prefer the typical stress pattern (N=1 or V=2) in novel English disyllables (Guion et al., 2003), show higher decision times and error rates (in a grammatical category assignment task) for atypical (N=2 or V=1) than for typical disyllables (Arciuli and Cupples, 2003), and produce stronger acoustic cues for typical stress in (real) English N/V pairs (Sereno and Jongman, 1995).[12] It is also known that for English disyllables, word stress is misperceived more often as initial in "trochaic-biasing" contexts, where the preceding syllable is weak or the following syllable is heavy; and more often as final in analogously "iambic-biasing" contexts. This effect is more pronounced for nouns than for verbs; and nouns occur more frequently in trochaic contexts (Kelly and Bock, 1988; Kelly, 1988, 1989). M. Kelly and collaborators have argued these facts are responsible for both the N/V stress asymmetry and the directionality of N/V pair stress shifts.

[11] A standard example is final obstruent devoicing, a common change cross-linguistically. Blevins (2006) summarizes the evidence that there are several articulatory and perceptual reasons why final voiced obstruents could be heard as unvoiced, but no motivation for the reverse process (final unvoiced obstruents heard as voiced).

[12] For example, Sereno and Jongman find that the ratio of amplitudes of the first and second syllables —an important cue to stress—is greater for initially-stressed N/V pairs (e.g. *police*) read in noun context, compared to verb context.

3.2 Frequency

Stress shift in English N/V pairs—in particular the most common change, the *diatonic stress shift* (DSS; $\{2,2\}\rightarrow\{1,2\}$)—has been argued to be a case of analogical change (Hock, 1991; Kiparsky, 1995) or lexical diffusion (Sherman, 1975; Phillips, 1984, 2006); indeed, the relationship between the two is controversial (see Phillips, 2006, vs. Kiparsky, 1995; Janda and Joseph, 2003). For both types of change, frequency has been argued to play a role in determining which forms change first; in particular, lower-frequency forms are said to be more susceptible to analogical change (e.g. Mańczak, 1980), or to change first in cases of lexical diffusion which require "lexical analysis" (Phillips, 2006), such as N/V stress shifts. This type of effect has been demonstrated for the most common N/V stress shift: words with lower frequencies are more likely to undergo the DSS (Phillips, 1984; Sonderegger, 2010). More precisely, among a set of N/V pairs pronounced as $\{2,2\}$ in 1700, those with lower present-day, combined (N+V) frequency are more likely to have changed to $\{1,2\}$ by the present.[13]

There is, however, an important ambiguity to this finding: present-day frequencies are used, under the implicit assumption that they have changed little diachronically. We must therefore distinguish between (at least) two hypotheses for why low-frequency words change (on average) earlier:

1. Words' relative frequencies stay approximately constant diachronically. In a given year, word a is more likely than word b to change if a is less frequent than b.
2. A word changes when its frequency drops below a (possibly word-specific) critical value.

Under Hypothesis 2, the reason present-day frequencies are on average lower for words which have changed is that their frequencies have decreased diachronically.

We can begin to differentiate between these hypotheses by examining diachronic frequency trajectories for N/V pairs which have changed, and checking whether they show negative trends. Real-time frequency trajectories (combined N+V frequencies) were found for 6 N/V pairs (*combat*, *decrease*, *dictate*, *perfume*, *progress*, *protest*) which have shifted stress since 1700.[14] Fig. 4 shows

[13] Sonderegger (2010) argues that frequency and phonological structure interact to influence which words undergo the DSS first. Here we refer to the finding that there is a significant main effect of frequency once prefix class is taken into account.

[14] A reviewer suggests that either N or V frequency alone would be a more relevant measure for particular changes, i.e. change in the stress of the N form might be triggered by change in its frequency or in the V form's frequency. This seems plausible, and we plan to consider frequency trajectories more carefully in future work; here we consider N+V frequency rather than N or V frequency alone for compatibility with previous work (Phillips, 1984; Sonderegger, 2010), where N+V frequency is used.

frequency trajectories alongside pronunciation trajectories for these pairs.

Frequencies were found by sampling from prose written by British authors in the Literature Online (LiOn) database, then normalizing against frequency trajectories for a set of 4 reference words. Details and some justification for this normalization step are given in Appendix C.[15]

All words show negative correlations between year and N+V frequency, 4/6 of which are significant ($p < 0.05$).[16] Although any conclusion must be tentative in view of the small number of frequency trajectories considered, these negative correlations lend support to Hypothesis 2, and rule out the hypothesis that the frequency trajectories for N/V pairs show no long-term trends. We thus adopt the working hypothesis that change occurs in an N/V pair when its frequency drops below a critical level.

3.3 Analogy/Coupling

A very broad explanation often invoked in language change is *analogy*: linguistic elements which are similar by some criterion change to become more similar. In the case of N/V pairs, it has been suggested that the most common stress shift, from {2,2} to {1,2}, could be due to analogical pressure: given the strong tendency in English for nouns to have earlier stress than verbs (e.g. Ross, 1973), speakers "regularize" {2,2} pairs to follow the dominant pattern of stress in the lexicon (Phillips, 2006:37–9).

In the context of our N/V diachronic pronunciation trajectories, we restate analogy as *coupling* between trajectories. We can check for coupling effects at two levels: within N/V pairs, and within prefix classes.

Within N/V pairs We have shown that to a first approximation, trajectories move along the $\{1,1\} \leftrightarrow \{1,2\} \leftrightarrow \{2,2\}$ axis (only one of the N or V forms changes at a time), and the pronunciation {2,1} never occurs. These facts are strong evidence for coupling between the N and V forms of each pair: if there were no coupling, there would be no reason why {2,1} could not occur, since N=2 and V=1 do occur independently in the dataset. There would also be no reason for trajectories to mostly move along this axis.

[15] `lion.chadwyck.com`. Only 6 words/4 reference words were considered because finding trajectories is time-intensive.

[16] Alphabetically: $r = -0.78$ ($p < 0.001$), $r = -0.78$ ($p < 0.1$), $r = -0.79$ ($p < 0.01$), $r = -0.32$ ($p > 0.25$), $r = -0.76$ ($p < 0.05$), $r = -0.74$ ($p < 0.01$)

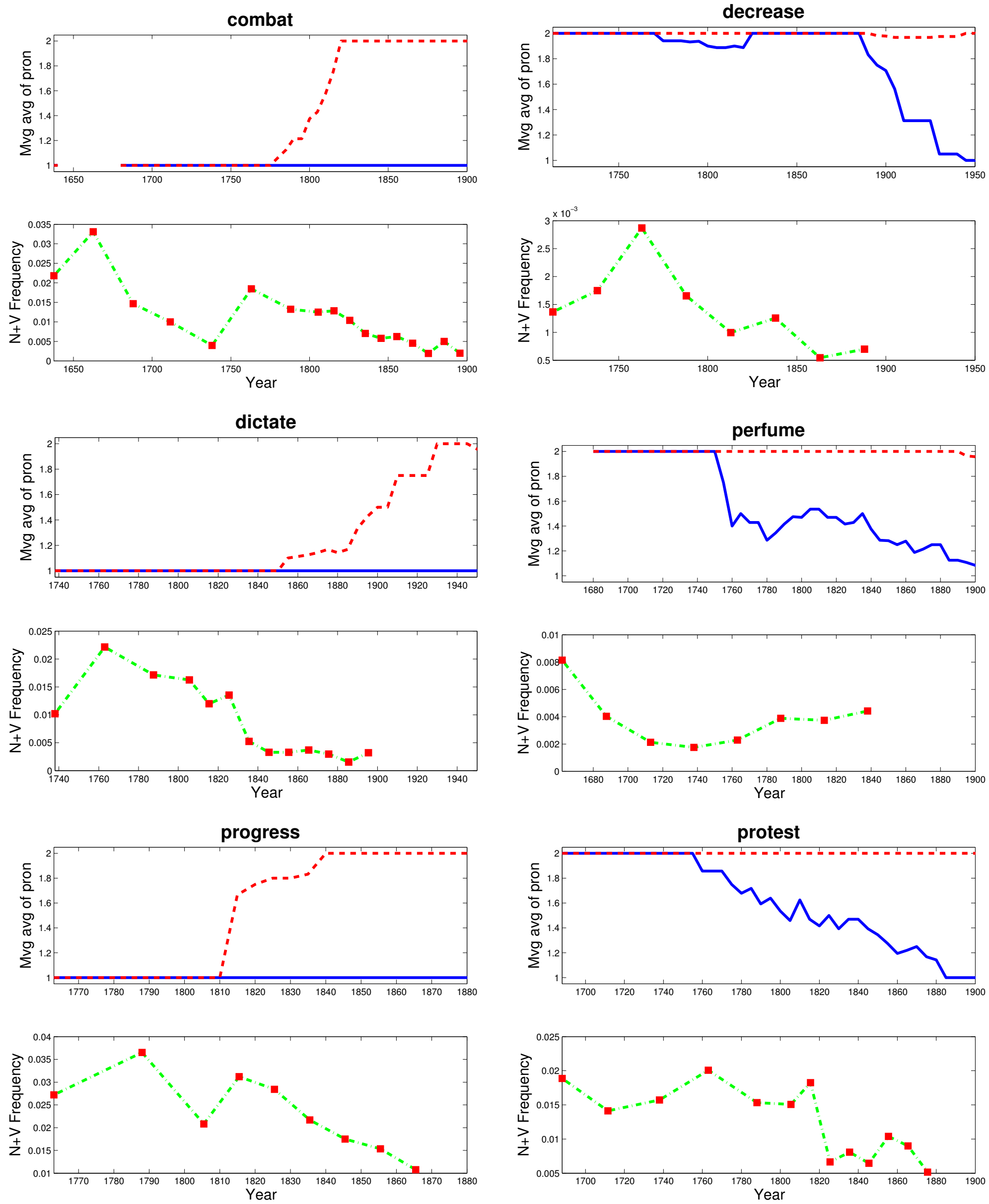


Figure 4: Frequency (bottom) and pronunciation (top) trajectories for *combat*, *decrease*, *dictate*, *perfume*, *progress*, *protest*.

Within prefix classes Impressionistically, over all N/V pair trajectories, those for pairs sharing a prefix often seem more similar than would be expected by chance. For example, many *re-* pairs were historically {2,2}, then began to change sometime 1875–1950. We would like a principled way to test the hypothesis of coupling between the trajectories for words in the same prefix class; to do so, we need a way to test how much two words "change like" each other, or how similar their trajectories are. We use a simple distance metric over trajectories' dissimilarity ("distance"), denoted $d(w, w')$ (for N/V pairs w and w').[17]

Finding $d(w, w')$ for all possible word pairs defines a graph $G(d)$ with nodes $w_1, \ldots, w_{149}$, and edges $d(w_i, w_j)$ equal to the distance between w_i and w_j's trajectories. This structure suggests a way of testing whether, given a group of words which are linguistically related, their trajectories are similar: check the goodness of the cluster formed by their vertices in G. For a subset of vertices $C \in [n]$ of $G = (V, E)$, define $R(C)$ to be the mean in-degree of C minus the mean out-degree of C.[18] $R(C)$ will be high if most vertices of C are on average closer to each other than to vertices in $V \setminus C$. This quantity is adapted from a common metric for finding community structure in networks (Newman and Girvan, 2004), with the important difference that here we are only evaluating *one* hypothesized community rather than a partitioning of G into communities.

As a measure of the goodness of a cluster C, let $p(C) \in [0, 1]$ be the empirical p-value, defined as the location of $R(C)$ on the distribution of R for all communities of size $|C|$ in G. The closer the value of $p(C)$ to 0, the more similar the trajectories for words in C are, compared to trajectories of a random set of words of size $|C|$. This setup can be used to test whether words in List 1 which share a prefix have similar trajectories. Table 5 shows $p(C)$ for all prefix classes of size $|C| > 2$.

Many potential prefix classes have small $p(C)$, confirming the initial intuition that N/V pairs sharing a prefix tend to have more similar trajectories. The *com-*/*con-* and *im-*/*in-* categories are particularly interesting because they suggest that it is a shared morphological prefix rather than simply shared initial segments which correlates with trajectory similarity. $p(C)$ for combined *com-* and *con-* is lower than either alone, and the same holds for *im-*/*in-*; this makes sense under the assumption that in- and *im-* are allophones of a single underlying prefix.

We also find that larger classes have lower $p(C)$: there is a significant negative relationship between $|C|$ and $\log(p(C))$ ($r = -0.72$, $p < 10^{-4}$) for the data in Table 5. That is, larger classes show stronger analogical effects, in the sense of trajectory similarity considered here.

[17] Over both N and V trajectories, the sum of the mean trajectory difference and the mean difference between trajectory first differences. Details are given in (Sonderegger, 2009).

[18] $R(C) = \dfrac{\sum\limits_{(i,j):i,j \in C} d(i,j)}{\binom{|C|}{2}} - \dfrac{\sum\limits_{(i,j):i \in C, j \notin C} d(i,j)}{\binom{n}{2} - \binom{|C|}{2} - \binom{n-|C|}{2}}$

C	$\|C\|$	$p(C)$	C	$\|C\|$	$p(C)$	C	$\|C\|$	$p(C)$
a-	10	0.270	dis-	5	0.746	re-	24	0.011
com-	5	0.067	ex-	6	0.981	re- (bound)	8	0.576
comp-	3	0.032	im-	4	0.021	re- (unbound)	16	0.0017
con-	17	0.001	in-	12	0.029	sub-	3	0.710
cont-	4	0.266	im-/in-	16	0.004	sur-	2	0.475
conv-	4	0.033	out-	10	0.055	trans-	3	0.173
com-/con-	22	0.0005	per	3	0.263	up-	7	0.196
de-	7	0.285	pre	5	0.065			
de- w/o des-	5	0.050	pro	4	0.078			

Table 5: Prefix class $p(C)$ values, $|C| > 2$. "bound"=re-μ, where μ is a bound morpheme.

4 Modeling

We have so far described the diachronic dynamics of variation and change in the stress of N/V pairs, and proposed causes for these dynamics. We now build dynamical systems models to test whether some proposed causes, implemented in the learning algorithm used by individuals, lead to one aspect of the observed dynamics at the population level: change following long-term stability, which in the language of dynamical systems corresponds to the presence of a bifurcation. This is only one of the multiple patterns observed in the data; the remainder are in part addressed elsewhere (Sonderegger, 2009; Sonderegger and Niyogi, 2010) and in part left to future work.

4.1 The dynamical systems approach

We derive models in the dynamical systems framework, which over the past 15 years has been used to model the interaction between language learning and language change in a variety of settings (Niyogi and Berwick, 1995, 1996; Komarova et al., 2001; Yang, 2001, 2002; Mitchener, 2005; Niyogi, 2006; Pearl and Weinberg, 2007) This framework is not a theory of language change, but a formalism to test theories of how change occurs. More precisely, it allows us to determine the diachronic, population-level consequences of assumptions about the learning algorithm used by individuals, as well as assumptions about population structure, the input received by learners, etc.

Our models are *discrete dynamical systems*, or *iterated maps*.[19] Given a domain X, an iterated

[19] See (Strogatz, 1994) for an introduction to dynamical systems.

map is a function $f : X \to X$ that "iterates" the system by one step. If a system has value $\alpha_t \in X$ at step t, it has value $\alpha_{t+1} = f(\alpha_t) \in X$ at step $t+1$. In models considered here, $X = [0,1]$. For example, $\alpha_t \in [0,1]$ will mean that at time t, the probability a random N example from the population (for a particular N/V pair) is produced with final stress is $P(N=2) = \alpha_t$, and the probability it is produced with initial stress is $P(N=1) = 1-\alpha_t$. Because we have assumed no coupling between the N and V forms, the situation for V, represented for example by $\beta_t \in [0,1]$, would be the same.[20]

Example 4.1. Let $I = [0,1]$, and let $f(x) = x^a$, where $a > 0$, so that $\alpha_{t+1} = \alpha_t^a$.

Solving for α_t gives $\alpha_t = {\alpha_0}^{at}$. However, unlike in this example, for a given f it is usually impossible to explicitly solve for α_t as a function of the initial state α_0. The dynamical systems viewpoint is to instead look at the system's long-term behavior as a function of the initial state.

Definition 1. $\alpha_* \in X$ is a *fixed point* of f if $\alpha_* = f(\alpha_*)$.

In the example, 0 and 1 are fixed points. However, when a is fixed, there is a qualitative difference between them. For a fixed $0 < a < 1$, for any initial state $\alpha_0 \neq 0$, $\lim\limits_{t\to\infty} \alpha_t = 1$. 0 is "unstable" in the sense that perturbing α_0 from 0 gives different long-term behavior ($t \to \infty$), while 1 is "stable" in the sense that perturbing α_0 from 1 does not.

Definition 2. A fixed point α_* is *stable* if $\lim\limits_{t\to\infty} \alpha_t = \alpha_*$ for α_0 near α_*, and *unstable* otherwise.

Stability turns out to be equivalent to a simple condition on f: a fixed point α_* is stable if and only if $|f'(\alpha_*)| < 1$, where f' denotes the derivative of f.

Definition 3. A *bifurcation* occurs when the number or stability of fixed points changes as a system parameter is changed.

For example, in Ex. 4.1, there is a bifurcation at $a = 1$ where the fixed points 0 and 1 exchange stabilities.

A central insight of the dynamical systems approach to modeling language change is that the pattern characterizing much language change—the sudden onset of change following a long period of stability—can be understood as a bifurcation in which a fixed point loses stability as some system parameter drifts past a critical value (Niyogi, 2006). In linguistic populations, system

[20] In models of coupling between the N and V forms of a pair, the domain is $[0,1]^2$, with (α_t, β_t) corresponding to the N and V probabilities at t.

parameters could be the frequency of a word or cue, the probability of misperceiving one segment as another, or the relative frequency of contact with speakers of one dialect versus another.

Although we mostly do not give derivations here, the task of a dynamical systems analysis is determining how the location, number, and stability of fixed points vary as a function of system parameters.

We make the following assumptions in all models considered below:

- Learners in generation n learn from generation $n-1$.
- Each example a learner in generation n hears is equally likely to come from any member of generation $n-1$.
- Each generation has infinitely many members.
- Each learner receives an identical number of examples.

These are idealizations, adopted here to keep models relatively simple. The effects of dropping each assumption are explored in (Niyogi, 2006; Sonderegger, 2009).

We also assume here that probabilities of producing initial vs. final stress for nouns and verbs are learned separately: that is, there is no "coupling" between them. However, a range of models for the N/V case incorporating coupling are considered in (Sonderegger, 2009; Sonderegger and Niyogi, 2010).

4.2 Model 1: Probability matching, mistransmission

Consider a population of learners following the above assumptions. Member i of generation t learns a probability $p_{i,t} \in [0,1]$, which characterizes the probability with which she uses form 2, versus form 1. As input to learners of the next generation, she produces form 2 examples with probability $p_{i,t}$ and form 1 examples with probability $1 - p_{i,t}$.

Let α_t be the mean value of $p_{i,t}$. We add in mistransmission errors (§3.1) via *mistransmission probabilities* that one form is heard when the other was intended:

$$a = P(1 \text{ heard} \,|\, 2 \text{ intended}), \quad b = P(2 \text{ heard} \,|\, 1 \text{ intended})$$

In generation $t+1$, learner i sets $p_{i,t+1}$ by probability matching, as follows:

- Draw N examples from generation t. Let $k_{i,t+1}$ be the number of examples heard as form 2.
- Set $p_{i,t+1} = k_{i,t+1}/N$

The evolution equation in this case works out to

$$\alpha_{t+1} = f(\alpha_t) = \alpha_t(1-a) + (1-\alpha_t)b \tag{1}$$

(Sonderegger, 2009, §5.2.2). Solving $f(\alpha_*) = \alpha_*$, there is a unique, stable fixed point at $\alpha_* = \frac{b}{a+b}$. The location of α_* does not depend on N, meaning word-frequency plays no role in the dynamics.

The dynamics show no bifurcations: as system parameters a, b, N are varied, the fixed point's location changes smoothly as a function of the ratio of the mistransmission probabilities. Thus, this model does not show the desired property of change following long-term stability, as a system parameter passes a critical value.

4.3 Model 2: Discarding

Model 1 assumes that each learner hears N examples, every one of which is heard as form 1 or form 2. We now consider a population of learners where each example can be heard as form 1, form 2, or *discarded*. Learners then probability match based on only a subset of the data, the non-discarded examples.[21] For the case of N/V pair stress, the experimental literature suggests one (speculative) reason English learners might discard some examples. Suppose learners discard examples where they are uncertain about stress placement. Given that the acoustic cues to stress in typically-stressed examples (N=1, V=2) are stronger than in atypically-stressed examples (N=2, V=1) for at least some speakers (Sereno and Jongman, 1995), some atypically-stressed examples learners might be discarded by learners.

We define *discarding probabilities* that form 1 or form 2 examples are discarded:

$$r_1 = P(\text{discarded} \,|\, 1 \text{ intended}), \quad r_2 = P(\text{discarded} \,|\, 2 \text{ intended})$$

and define $p_{i,t}$ as above. For learner i in generation $t+1$, the algorithm is:

- Draw N examples from generation t, of which $k^{(2)}_{i,t+1}$ are heard as form 2, $k^{(1)}_{i,t+1}$ as form 1, and $N - k^{(1)}_{i,t+1} - k^{(2)}_{i,t+1}$ are discarded.
- Set $p_{i,t+1} = \begin{cases} r & : k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1} = 0 \\ \frac{k^{(2)}_{i,t+1}}{k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1}} & : k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1} \neq 0 \end{cases}$
  where $r \in [0,1]$.

[21] This type of learner is similar in spirit to the idea of "input filtering" suggested in Lisa Pearl's computational studies of English acquisition and change (Pearl, 2007, *et seq.*), where learners consider only examples relevant to the cue currently being set.

That is, the learner's default strategy when *all* examples are discarded is to set $p = r$ (for r fixed). For any N and non-zero discarding probabilities, there is always some chance (though possibly very small) that all examples are discarded. Where r comes from is left ambiguous; for example, it could be the percentage of known disyllabic words with final stress.

The evolution equation works out to

$$\alpha_{t+1} = \frac{\alpha_t(1-r_2)}{(1-r_1)+\alpha_t(r_1-r_2)} + (\alpha_t(r_1-r_2))^N[r - \frac{\alpha_t(1-r_2)}{(1-r_1)+\alpha_t(r_1-r_2)}] \tag{2}$$

(Sonderegger, 2009, §5.5.1). In the high-frequency ($N \to \infty$) limit, this reduces to:

$$\alpha_{t+1} = \frac{\alpha_t(1-r_2)}{(1-r_1)+\alpha_t(r_1-r_2)} \tag{3}$$

In practice, the long-term dynamics of Eqn. 3—in particular the location of the unique stable fixed point—are extremely similar to the true (frequency-dependent) evolution equation (Eqn. 2) for N greater than a small value (≈3–5, depending on the values of r_1 and r_2). That is, the long-term dynamics are only affected by frequency for very small N. We thus only consider Eqn. 3 here.

Solving $\alpha_{t+1} = \alpha_t$ in Eqn. 3 gives two fixed points: $\alpha_- = 0$ and $\alpha_+ = 1$. There is a bifurcation at $r_1 = r_2$: for $r_1 < r_2$, α_- is stable and α_+ is unstable; for $r_2 < r_1$, α_- is unstable and α_+ is stable. Intuitively, the form with a higher probability of being discarded is eliminated.

4.4 Model 3: Discarding+mistransmission

Consider now a simple model incorporating both mistransmission and discarding. For a given example, define $a, b, R \in [0,1]$ such that:

$$\begin{aligned} P(H=1 \mid I=1) &= 1-b, \quad P(H=2 \mid I=1) = bR, \quad P(\text{discarded} \mid I=1) = b(1-R) \\ P(H=2 \mid I=2) &= 1-a, \quad P(H=1 \mid I=2) = aR, \quad P(\text{discarded} \mid I=2) = a(1-R) \end{aligned}$$

where H="heard", I="intended". a and b are now the probabilities of not hearing a form i example as form i. When this occurs, the probability the example is heard as the wrong form, versus being discarded, is R.

The learning algorithm for member i of generation $t+1$ is the same as in Model 2, but now $k^{(2)}_{i,t+1}$ may include some mistransmitted form 1 examples (and similarly for $k^{(1)}_{i,t+1}$).

Analysis of the resulting evolution equation (Sonderegger, 2009, §5.5.2) shows there is a single fixed point, α_*, and thus no bifurcations. Similarly to Model 2, there is essentially no effect of

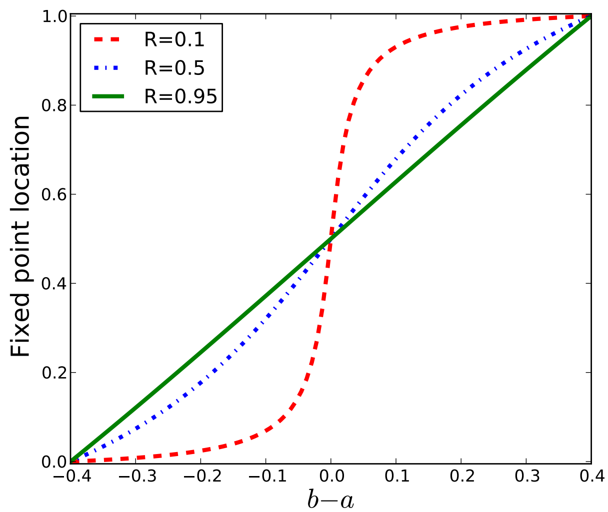


Figure 5: Location of α_* vs. $b-a$, $a+b=0.5$, for different values of R.

frequency on long-term dynamics for N above a relatively small value, and we thus consider the high-frequency limit of the evolution equation. The location of α_* as a function of $a-b$ as R is varied is plotted in Fig 5. R controls how "bifurcation-like" the curve is: for R small α_* changes rapidly at $a=b$; for $R\to 1$, α_* varies smoothly as a function of $a-b$. However, there is no bifurcation *to* bifurcation-like behavior: adding any mistransmission $R>0$ eliminates the bifurcation seen in Model 2.

5 Discussion

We have described a diachronic corpus of N/V pair stress, the dynamics of stress shifts observed in the corpus, and several proposed factors driving this change: mistransmission, word frequency, and analogy. We then determined the population-level, diachronic dynamics of three models of learning, to explore which models show bifurcations, i.e. which give stability followed by sudden change as system parameters are varied. We did not evaluate models with respect to the frequency or analogical effects observed in the corpus (§3.2–3.3); however, both are considered in the larger set of models described elsewhere (Sonderegger, 2009; Sonderegger and Niyogi, 2010).

Following an idea proposed by Niyogi (2006), we suggest that bifurcations in the diachronic dynamics of linguistics population are a possible explanation for the core of the actuation problem: how and why does language change begin in a community, following long-term stability? This viewpoint suggests a powerful test of theories of the causes of language change: do their diachronic dynamics show bifurcations? We found that mistransmission alone (Model 1) does not give bifurcations, while discarding alone (Model 2) does: the form more likely to be discarded is eliminated from the population. Combining mistransmission and discarding (Model 3) eliminates bifurcations, but gives more or less bifurcation-like behavior as the relative probability of mistransmission and discarding is varied.

In line with other computational work on population-level change where several models are considered (e.g. Liberman, 2000; Daland et al., 2007; Baker, 2008; Troutman et al., 2008), the different dynamics of Models 1–3 illustrate that different proposed causes for change at the individual level, each of which seems plausible a priori, can have very different population-level diachronic outcomes. Among models tested here, only those including discarding showed bifurcations (Model 2) or bifurcation-like behavior (Model 3); the model including only mistransmission (Model 1) did not. Given the popularity of mistransmission-based explanations of phonological change, this result illustrates an important point: because of the non-trivial map between individual learning and population dynamics, population-level models are necessary to evaluate any theory of why language change occurs.

References

Arciuli, J. and Cupples, L. (2003). Effects of stress typicality during speeded grammatical classification. *Language and Speech*, 46(4):353–374.

Baayen, R., Piepenbrock, R., and Gulikers, L. (1996). *CELEX2 (CD-ROM)*. Linguistic Data Consortium, Philadelphia.

Baker, A. (2008). Addressing the actuation problem with quantitative models of sound change. *Penn Working Papers in Linguistics*, 14(1):1–13.

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, 32(2):117–166.

Boyer, A. (1700). *The royal dictionary: in two parts...* London.

Daland, R., Sims, A., and Pierrehumbert, J. (2007). Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th Annual Meeting of the Associ-*

ation of Computational Linguistics, pages 936–943. Association for Computational Linguistics, Prague, Czech Republic.

Guion, S., Clark, J., Harada, T., and Wayland, R. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Language and Speech*, 46(4):403–427.

Hansson, G. (2008). Diachronic explanations of sound patterns. *Language & Linguistics Compass*, 2:859–893.

Hock, H. (1991). *Principles of Historical Linguistics*. Mouton de Gruyter, Berlin, 2nd edition.

James, W. and Molé, A. (1847). *Dictionary of the English and French languages...* Leipzig.

Janda, R. and Joseph, B. (2003). On language, change, and language change—or, of history, linguistics, and historical linguistics. In Joseph, B. and Janda, R., editors, *Handbook of Historical Linguistics*, pages 3–180. Blackwell, Oxford.

Kelly, M. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition*, 30:107–137.

Kelly, M. (1989). Rhythm and language change in English. *Journal of Memory & Language*, 28:690–710.

Kelly, M. and Bock, J. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance*, 14(3):389–403.

Kiparsky, P. (1995). The phonological basis of sound change. In Goldsmith, J., editor, *The Handbook of Phonological Theory*, pages 640–670. Blackwell, Cambridge, MA.

Komarova, N., Niyogi, P., and Nowak, M. (2001). The evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1):43–60.

Liberman, M. (2000). The 'lexical contract': Modeling the emergence of word pronunciations. Ms., University of Pennsylvania.

Mańczak, W. (1980). Laws of analogy. In Fisiak, J., editor, *Historical Morphology*, volume 17 of *Trends in Linguistics. Studies and Monographs*, pages 283–288. Mouton, The Hague.

Mitchener, W. G. (2005). Simulating language change in the presence of non-idealized syntax. In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition*, pages 10–19, Ann Arbor, Michigan. Association for Computational Linguistics.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113.

Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. MIT Press, Cambridge.

Niyogi, P. and Berwick, R. (1995). The logical problem of language change. AI Memo 1516, MIT.

Niyogi, P. and Berwick, R. (1996). A language learning model for finite parameter spaces. *Cognition*, 61(1-2):161–193.

Ohala, J. (1981). The listener as a source of sound change. In Masek, C., Hendrick, R., and Miller, M., editors, *Papers from the Parasession on Language and Behavior*, pages 178–203. Chicago Linguistic Society, Chicago.

Pearl, L. (2007). *Necessary Bias in Language Learning*. PhD thesis, University of Maryland.

Pearl, L. and Weinberg, A. (2007). Input filtering in syntactic acquisition: Answers from language change modeling. *Language Learning and Development*, 3(1):43–72.

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

Phillips, B. (2006). *Word Frequency and Lexical Diffusion*. New York: Palgrave.

Ross, J. (1973). Leftward, ho! In Anderson, S. and Kiparsky, P., editors, *Festschrift for Morris Halle*, pages 166–173. Holt, Rinehart and Winston, New York.

Sereno, J. and Jongman, A. (1995). Acoustic correlates of grammatical class. *Language and Speech*, 38(1):57.

Sherman, D. (1975). Noun-verb stress alternation: An example of the lexical diffusion of sound change in English. *Linguistics*, 159:43–71.

Sonderegger, M. (2009). Dynamical systems models of language variation and change: An application to an English stress shift. Masters paper, Department of Computer Science, University of Chicago.

Sonderegger, M. (2010). Testing for frequency and structural effects in an English stress shift. In Cleary-Kemp, J., Cohen, C., Farmer, S., Kassner, L., Sylak, J., and Woodley, M., editors, *Proceedings of the Berkeley Linguistics Society 36*.

Sonderegger, M. and Niyogi, P. (2010). Combining data and mathematical models of language change. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1019–1029, Uppsala, Sweden. Association for Computational Linguistics.

Strogatz, S. (1994). *Nonlinear Dynamics and Chaos*. Addison-Wesley, Reading, MA.

Troutman, C., Goldrick, M., and Clark, B. (2008). Social networks and intraspeaker variation during periods of language change. *Penn Working Papers in Linguistics*, 14(1):325–338.

Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory of language change. In Lehmann, W. and Malkiel, Y., editors, *Directions for Historical Linguistics*, pages 95–188. University of Texas Press, Austin.

Yang, C. (2001). Internal and external forces in language change. *Language Variation and Change*,

12(3):231–250.
Yang, C. (2002). *Knowledge and Learning in Natural Language*. Oxford University Press.

Appendix A: Word list from Sherman (1975) (List 1)

Script indicates first reported pronunciation: {1,1}, **{2,2}**, *{1,2}*

abstract	*compact*	**contrast**	**discharge**	*impact*	*invert*	**permit**	*rebel*	**rehash**	*torment*
accent	*compound*	*converse*	discord	**import**	*legate*	*pervert*	**rebound**	**reject**	transfer
addict	**compress**	*convert*	**discount**	**impress**	**misprint**	postdate	**recall**	*relapse*	**transplant**
address	**concert**	*convict*	**discourse**	*imprint*	*object*	**prefix**	**recast**	**relay**	*transport*
affect	*concrete*	*convoy*	*egress*	*incense*	*outcast*	**prelude**	**recess**	**repeat**	*transverse*
affix	*conduct*	**decoy**	**eject**	**incline**	*outcry*	*premise*	**recoil**	*reprint*	traverse
alloy	*confect*	**decrease**	**escort**	**increase**	*outgo*	presage	*record*	**research**	**undress**
ally	*confine*	**defect**	*essay*	**indent**	*outlaw*	*present*	**recount**	**reset**	*upcast*
annex	*conflict*	**defile**	**excerpt**	*infix*	**outleap**	*produce*	**redraft**	sojourn	*upgrade*
assay	*conscript*	**descant**	**excise**	*inflow*	*outlook*	progress	**redress**	*subject*	*uplift*
bombard	*conserve*	*desert*	*exile*	**inlay**	*outpour*	*project*	*refill*	sublease	upright
cement	*consort*	**detail**	**exploit**	*inlet*	*outspread*	**protest**	**refit**	**sublet**	**uprise**
collect	**content**	dictate	*export*	*insert*	*outstretch*	purport	**refund**	**surcharge**	*uprush*
combat	*contest*	**digest**	**extract**	*inset*	*outwork*	**rampage**	**refuse**	**survey**	*upset*
commune	**contract**	**discard**	*ferment*	*insult*	**perfume**	**rebate**	**regress**	**suspect**	

Appendix B: Sample of words in use 1700–2007 (List 2)

Script indicates pronunciation from (Boyer, 1700), as above. Asterisk indicates that 1700, 1847 (James and Molé, 1847), and 2007 (Cambridge Advanced Learner's Dictionary, OED) entries are not identical. Sample selection is described in §2.

abuse	bottom	*contest*	envy	harbour	measure	**proceed***	**repeal**	table	whistle
accent	breakfast	*contract*	*exile**	hollow	mention	**protest***	**repose**	tally	witness
advance	buckle	*convict*	**express**	**import***	merit	purchase	**reserve**	thunder	
affront	bundle	cover	favour	**increase***	motion	puzzle	**review**	title	
ally*	butter	**decrease***	ferret	interest	murder	quarry	rival	*torment*	
anchor	*cement**	**decree**	flourish	iron	muster	reason	saddle	travel	
arrest	challenge	diet	**forecast***	journey	order	**redress**	second	treble	
assault	channel	**digest***	forward	level	outlaw	**reform**	shiver	triumph	
assay	**command**	**dispatch**	gallop	levy	pepper	**regard**	shoulder	trouble	
attack	**concern**	**dissent**	glory	licence	plaster	**relapse***	squabble	value	
bellow	*conduct*	**distress**	hammer	license	*premise**	relish	stable	visit	
blunder	*consort*	double	handle	matter	*present*	**remark**	stomach	vomit	

Appendix C: Frequency trajectory normalization

Because LiOn only gives absolute counts, we normalized the N/V pair frequency trajectories in §3.2 by the (summed) frequency trajectories of four words from the Swadesh list: *red*, *walk*, *man*, *flower*. For the conclusion reached in §3.2 (that the N/V pairs considered decrease in frequency over time) to be valid, it must be the case that this set of reference words remains approximately constant in frequency over time. We checked that these words' frequencies show no time trends in two ways. First, when normalized by the LiOn frequency trajectory of one extremely frequent word (*the*), whose frequency presumably is approximately constant diachronically, the sum frequency of the reference words shows no time trend ($p > 0.1$ for both Pearson and Spearman correlations). Second, the summed *relative* frequencies of the set of reference words (i.e. occurrences per million) show no time trends ($p > 0.15$, Pearson & Spearman) in the Corpus of Historical American English (COHA), which includes 400 million words from 1810–2000.[22] Although COHA covers a different dialect of English and a somewhat different time period than the N/V pair frequency trajectories, it is the largest available diachronic corpus of English, and thus provides some reassurance that the summed frequency of the set of reference words chosen is not especially volatile.

[22] `corpus.byu.edu/coha/`, beta version.