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

Morgan Sonderegger*  Partha Niyogi†
morgan@cs.uchicago.edu  niyogi@cs.uchicago.edu

Draft of 1/2010

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

¹Why 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)
diachronic, population-level consequences of assumptions about the process of language learning by individuals (references omitted).

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 sudden 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. 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. N/V pairs are a productive class (YouTube, google). To get a rough count of N/V pairs, the overlap of the CMU Dictionary (American English) with the British National Corpus word-frequency list (Leech et al., 2001) gives 3185 pairs, of which 1783 have N or V frequency > 0, and 647 have both N and V frequency > 0.

Out of the four logically possible stress patterns, all current N/V pair pronunciations follow one of three patterns:

<table>
<thead>
<tr>
<th>N</th>
<th>V</th>
<th>Pronunciation</th>
</tr>
</thead>
<tbody>
<tr>
<td>{1, 1}</td>
<td>1σσ</td>
<td>anchor, fracture, forecast</td>
</tr>
<tr>
<td>{1, 2}</td>
<td>1σσ</td>
<td>consort, protest, refuse</td>
</tr>
<tr>
<td>{2, 2}</td>
<td>2σσ</td>
<td>cement, police, review</td>
</tr>
</tbody>
</table>

These models are sampled from a larger project (Sonderegger, 2009; Sonderegger and Niyogi, 2010a,b), 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.

The number of N/V pairs in English 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.

We abbreviate N=noun, V=verb, N/V=noun-verb throughout, and use curly brackets to denote N and V stress, where 1=initial stress, 2=final stress.
As discussed below, the \( \{2, 1\} \) pattern is also never observed diachronically. At any given time, variation exists in the pronunciation of some N/V pairs, e.g. *research, address* in current-day American English.

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 \( \{1, 2\} \). 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 while many words were \( \{1, 2\} \) by 1800, those that were not must have shifted at some point by 1975. We will reexamine Sherman’s interpretation below after examining the dynamics of an expanded dataset.

**Stability of \( \{1, 1\}, \{2, 2\} \)** 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 (App. B) is a set of 110 N/V pairs chosen at random from all N/V pairs Words were randomly chosen 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?

---

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

6To be more sure they have been in use, we would need frequency counts for the N and V forms in 1700 and 1847.
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 American dictionaries, published 1800–2003. 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 1700. 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) 0 (not listed). At a given time, the N or V forms for many words in List 1 are rare, archaic, or not yet in use. The combination N=2, V=1 is never reported.

Changes in individual N/V pairs’ pronunciations can be visualized by plotting the moving average of their N and V 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

\[
\begin{align*}
  f(1) &= 1, & f(2) &= 2, & f(1/2) &= f(2/1) = f(1.5) = 1.5
\end{align*}
\]

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

For a word \(w\) at time \(t\), the average of pronunciations reported in the time window \((t-25, t+25)\) (years) was plotted if \(\geq 2\) dictionaries 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 were used. A sample of the resulting 149 stress vs. time trajectories is given in Figs. 1–2.

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):

---

7The data and dictionary list are in (Sonderegger and Niyogi, 2010a); the data are available at http://people.cs.uchicago.edu/~morgan/??.

8If is often assumed that English words have a unique primary stress, so that “level stress” is not possible. Instead of discarding the 113 reports of level stress, we interpret them as equal preference for initial and final primary stress.

9In fact, dictionary authors often state that the first listed pronunciation is “primary”, meaning 1/2 really means 1 is heard more in the population than 2. In practice, 1/2 and 2/1 are uncommon enough that trajectories plotted using \(f(1/2) = 1.25, f(2/1) = 1.75\) look similar.

10All trajectories are given in (Sonderegger and Niyogi, 2010a), and posted at http://people.cs.uchicago.edu/~morgan/??.
Figure 1: Sample trajectories 1: change between endpoints.
Table 1: Distribution of data with both N and V stresses listed. “Var” means 1.5, 1/2, or 2/1.

<table>
<thead>
<tr>
<th>(pron_N, pron_V)</th>
<th>{1,1}</th>
<th>{1,var}</th>
<th>{1,2}</th>
<th>{var,2}</th>
<th>{2,2}</th>
<th>{var,var}</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of entries</td>
<td>7.1</td>
<td>4.6</td>
<td>57.1</td>
<td>7.1</td>
<td>21.8</td>
<td>2.2</td>
</tr>
</tbody>
</table>

1. \(\{2,2\} \rightarrow \{1,2\}\) (concert)  
2. \(\{1,1\} \rightarrow \{1,2\}\) (combat)  
3. \(\{1,2\} \rightarrow \{1,1\}\) (collect)  
4. \(\{1,2\} \rightarrow \{2,2\}\) (cement)  

A sample of each type is shown in Fig. 1.  
Change directly between \(\{1,1\}\) and \(\{2,2\}\) never occurs, suggesting that change occurs one form (N or V) at a time. This is borne out by the percentages of dictionary entries reporting variation in N, V, or neither, shown in Table 1. We see that while variation is reported in either N or V in 13.9% of entries which list both N and V pronunciations, variation in both N and V forms at once is reported in only 2.2% of such entries.  

What is the diachronic behavior of the observed variation? Examining all trajectories, we can make some impressionistic observations. Variation near endpoints (concert, Fig. 1(c)) is relatively common, but long-term variation away from endpoints (exile, Fig. 2(b)) is rare. Long-term variation in both N and V forms at once (rampage, Fig. 2(a)), is very rare (2 trajectories).  

\(\{2,1\}\) is never observed in the dataset, and we argue it is in fact “unstable” in the following
Figure 3: Schematic of observed changes. Self-loops indicate that \(\{1, 1\}, \{1, 2\}, \{2, 2\}\) are stable states. \(\{1, 0\}\) and \(\{0, 2\}\) are disyllabic words without V and N forms (respectively). Line thicknesses indicate transitions’ relative frequencies.

Entries “near” \(\{2,1\}\), such as \((N=2/1,V=1/2)\) are very rare (9 instances), and are scattered across different words and dictionaries. This means that the few times a trajectory drifts towards the region \(\text{pron}_N > \text{pron}_V\), it quickly moves away. In the language of dynamical systems (§4.1), this suggests the region contains an unstable fixed point (one which repels trajectories), \(\{2,1\}\).

We can summarize the observed diachronic facts as follows:

1. \(\{1,1\}, \{1,2\}, \{2,2\}\) are “stable states”, but variation around them often occurs. Long-term variation away from these states is rare.
2. Variation usually occurs in only one of N or V at a time.
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 accurately 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 know whether pronunciation variation exists in individuals or not: do individuals learn \textit{gradient} (a probability $\alpha \in [0, 1]$ of using one form versus another) or \textit{categorical} (each speaker uses one form exclusively) forms? We call these options \textit{within-speaker} and \textit{between-speaker} variation.\footnote{The terminology is slightly misleading because the structure of variation (the $\alpha$ stored) differs between speakers in “within-speaker” variation as well.}

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 following table lists the number of speakers (31 total, 18 male) who pronounced the noun form of \textit{research}, \textit{address}, or \textit{perfume}, exclusively with initial stress, exclusively with final stress, or used both:\footnote{See (Sonderegger and Niyogi, 2010a) for details, including the list of stories.}

<table>
<thead>
<tr>
<th>Word</th>
<th># N=1</th>
<th># Var</th>
<th># N=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>research</td>
<td>9</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>perfume</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>address</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Each speaker listed for a word used it at least 5 times.

Within-speaker variation thus does occur for N/V pairs, at least in this relatively small dataset. Nonetheless, this finding has important consequences for modeling. As has been pointed out in both dynamical systems (reference omitted) 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 suggests learners cannot simply be 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.
3 Motivations for change

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

3.1 Mistransmission

An influential line of research holds that 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 $\alpha$ as $\beta$, but rarely mishear $\beta$ as $\alpha$.\(^{13}\)

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 (e.g. Blevins and Garrett, 1998; Kavitskaya, 2001; Blevins, 2004; Barnes, 2006).

In the domain of N/V pair stress, 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), and 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). 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, 1988; Kelly and Bock, 1988; Kelly, 1989, et seq.). 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.

\(^{13}\)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).
3.2 Frequency

The possibility that sound change proceeds faster for some words than others, contrary to the Neogrammariann hypothesis, has remained controversial since it was proposed by Schuchardt (1885). More recent work argues for a relationship between word frequency and sound change in a variety of cases. Proponents argue that in some changes (e.g. English yod-dropping: Phillips (1981)), low-frequency words change first (on average); in others (e.g. English schwa-deletion: Hooper (1976)) high-frequency words change first.

However, frequency is commonly invoked in other types of changes, such as analogy and paradigmatic change. Because the N/V stress shift seems to fall into these categories rather than a classical sound change, it would not be surprising to find that word frequency plays a role in which N/V pairs shift stress. Phillips (1984) claims that the N/V stress shift is a case of “low-frequency first” change, but does not control for how long pairs have been in use. However, analysis of a controlled dataset (see Sonderegger, 2010) confirms Phillips’ hypothesis: among a set of N/V pairs pronounced as {2,2} in 1700 and in use up to the present, those with lower present-day, combined (N+V) frequency are more likely to have changed to {1,2}.

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 given 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 were found for 6 N/V pairs (combat, decrease, dictate, perfume, progress, protest) which have shifted stress since 1700. Fig. 4 shows frequency trajectories alongside pronunciation trajectories for these pairs.

Figure 4: Frequency (bottom) and pronunciation (top) trajectories for *combat*, *decrease*, *dictate*, *perfume*, *progress*, *protest*.
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 reference words. All words show negative correlations between year and N+V frequency, 5/6 of which are significant. Although these correlations are weak, they lend support to Hypothesis 2, and rule out the assumption 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 N/V context, we restate analogy as coupling between pronunciation 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 \( \{1,1\} \rightarrow \{1,2\} \rightarrow \{2,2\} \) 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 the \( \{1,1\} \rightarrow \{1,2\} \rightarrow \{2,2\} \) axis.

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; for this, 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”), \( d(w, w') \), for N/V pairs \( w \) and \( w' \).

---

15 lion.chadwyck.com. Only 6 words were considered because finding trajectories is time-intensive.
16 Alphabetically: \( r = -0.78 \) (\( p < 0.001 \)), \( r = -0.78 \) (\( p < 0.1 \)), \( r = -0.79 \) (\( p < 0.01 \)), \( r = -0.32 \) (\( p > 0.25 \)), \( r = -0.76 \) (\( p < 0.05 \)), \( r = -0.74 \) (\( p < 0.01 \))
17 Over both N and V trajectories, the sum of the mean trajectory difference and the mean difference between trajectory first differences (details in Sonderegger and Niyogi, 2010a)
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 $C$. $R(C)$ 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 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 2 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-*-, *im-*/*in-*-, and *re-* 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 (*in-*). For *re-* the difference between the productive (before bound morphemes) and unproductive (before unbound) versions is clearly reflected in $p(C)$.

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 2. This relationship is interesting given the main criticism of explanations invoking analogy, its arbitrariness: what determines whether a given set of forms which are somehow similar become more similar? In the current case, we can answer that analogy is more likely to affect larger sets of similar words. This makes intuitive sense: if analogy can affect any two or three related words as much as large classes, we would expect much less regularity in language change than is observed.

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

$^a$“bound”: re-$\mu$, where $\mu$ is a bound morpheme

Table 2: Prefix class $p(C)$ values, $|C| > 2$. 

14
4 Modeling

We have so far described (1) the diachronic dynamics of variation and change in the stress of N/V pairs (2) proposed causes for these dynamics. We now build dynamical systems models to test which proposed causes, implemented in the learning algorithm used by individuals, lead to one aspect of the observed dynamics at the population level: sudden 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; we leave the remainder 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 (Komarova et al., 2001; Yang, 2001, 2002; Mitchener, 2005; 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. Given a domain $X$, an iterated 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.

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 $\alpha_t$ gives $\alpha_t = \alpha_0^{a\cdot t}$. However, unlike in this example, for a given $f$ it is usually impossible to explicitly solve for $\alpha_t$ as a function of the initial state $\alpha_0$. The dynamical systems viewpoint is to instead look at the system’s long-term behavior as a function of the initial state.

---

19Strogatz (1994) is a readable introduction.
20In models of coupling between the N and V forms of a pair, which we will not have space to discuss here (see Sonderegger and Niyogi, 2010b), the domain is $[0, 1]^2$, with $(\alpha_t, \beta_t)$ corresponding to the N and V probabilities at $t$.  

15
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 \( \alpha_0 \neq 0 \), \( \lim_{t \to \infty} \alpha_t = 1 \). 0 is “unstable” in the sense that perturbing \( \alpha_0 \) from 0 gives different long-term behavior \( (t \to \infty) \), while 1 is “stable” in the sense that perturbing \( \alpha_0 \) from 1 does not.

Definition 2. A fixed point \( \alpha^* \) is stable if \( \lim_{t \to \infty} \alpha_t = \alpha^* \) for \( \alpha_0 \) near \( \alpha^* \), and unstable otherwise.

What makes dynamical systems analysis possible is that stability is equivalent to a simple condition on \( f \): a fixed point \( \alpha^* \) 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 – sudden 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. In linguistic populations, system 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:\(^{21}\)

- 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.

---

\(^{21}\)These are idealizations, adopted here to keep models simple enough to analyze. The effects of dropping these assumptions are explored in (Niyogi, 2006; Sonderegger and Niyogi, 2010b).
- Probabilities of producing initial vs. final stress for nouns and verbs are learned separately: there is no “coupling” between them.

### 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 \( \alpha_t \) be the mean value of \( p_{i,t} \). We add in mistransmission errors (§3.1) via \textit{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 \( \alpha_{t+1} = f(\alpha_t) = \alpha_t (1 - a) + (1 - \alpha_t) b \). Solving \( f(\alpha_*) = \alpha_* \), there is a unique, stable fixed point at \( \alpha_* = \frac{b}{a+b} \).

The dynamics thus 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. The location of \( \alpha_* \) also does not depend on \( N \), meaning word-frequency plays no role in the dynamics.

### 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 \textit{discarded}. Learners then probability match based on only a subset of the data, the non-discarded examples. This type of learner is one implementation of the idea of “input filtering” suggested in Lisa Pearl’s computational studies of English acquisition and change (Pearl, 2007; Pearl and Weinberg, 2007; Pearl, 2008): learners consider only examples relevant to the cue currently being set.
We define discarding probabilities that form 1 or form 2 examples are discarded:

\[ r_1 = P(\text{discarded} \mid \text{1 intended}), \quad r_2 = P(\text{discarded} \mid \text{2 intended}) \]

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 \( p_{i,t+1} = \begin{cases} \begin{align*} \frac{r}{k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1}} : k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1} &= 0 \\ \frac{k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1}}{k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1}} : k^{(1)}_{i,t+1} + k^{(2)}_{i,t+1} &\neq 0 \end{align*} \end{cases} \)

where \( r \in [0, 1] \).

The learner’s default strategy when all examples are discarded is to set \( p = r \) (for \( r \) fixed).

For large \( N \), the evolution equation works out to

\[ \alpha_{t+1} = \frac{\alpha_t (1 - r_2)}{(1 - r_1) + \alpha_t (r_1 - r_2)} \]

Solving \( \alpha_{t+1} = \alpha_t \) gives two fixed points: \( \alpha_- = 0 \) and \( \alpha_+ = 1 \). There is a bifurcation at \( r_1 = r_2 \); for \( r_1 < r_2 \), \( \alpha_- \) is stable and \( \alpha_+ \) is unstable; for \( r_2 < r_1 \), \( \alpha_- \) is unstable and \( \alpha_+ \) 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:

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

where \( H=\text{“heard”}, I=\text{“intended”}. \) \( a \) 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 \( k_{i,t+1}^{(2)} \) may include some mistransmitted form 1 examples (and similarly for \( k_{i,t+1}^{(1)} \)).
Analysis of the resulting evolution equation shows there is a single fixed point, \( \alpha_\ast \), and thus no bifurcations. The location of \( \alpha_\ast \) 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 \( \alpha_\ast \) changes rapidly at \( a=b \); for \( R \to 1 \), \( \alpha_\ast \) 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 suggest that bifurcations in the diachronic dynamics of linguistics population explain the

---

22For \( N \) large but finite, \( \alpha_- \) and \( \alpha_+ \) move off 0 and 1, but the dynamics otherwise remain the same.
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 recent computational models of population-level change (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 (as opposed to mistransmission) showed bifurcations. Given the popularity of mistransmission-based explanations of sound 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


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

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

<table>
<thead>
<tr>
<th>abstract</th>
<th>compact</th>
<th>contrast</th>
<th>discharge</th>
<th>impact</th>
<th>invert</th>
<th>permit</th>
<th>rebel</th>
<th>rehash</th>
<th>torment</th>
</tr>
</thead>
<tbody>
<tr>
<td>accent</td>
<td>compound</td>
<td>converse</td>
<td>discord</td>
<td>import</td>
<td>legate</td>
<td>pervert</td>
<td>rebound</td>
<td>reject</td>
<td>transfer</td>
</tr>
<tr>
<td>addict</td>
<td>compress</td>
<td>convert</td>
<td>discount</td>
<td>impress</td>
<td>misprint</td>
<td>postdate</td>
<td>recast</td>
<td>relapse</td>
<td>transplant</td>
</tr>
<tr>
<td>address</td>
<td>concert</td>
<td>convict</td>
<td>discourse</td>
<td>imprint</td>
<td>object</td>
<td>prefix</td>
<td>prefix</td>
<td>recall</td>
<td>transpose</td>
</tr>
<tr>
<td>affect</td>
<td>concrete</td>
<td>convoy</td>
<td>egress</td>
<td>incense</td>
<td>outcast</td>
<td>prelude</td>
<td>recess</td>
<td>receive</td>
<td>traverse</td>
</tr>
<tr>
<td>affix</td>
<td>conduct</td>
<td>decay</td>
<td>eject</td>
<td>incline</td>
<td>outcry</td>
<td>premise</td>
<td>recoil</td>
<td>record</td>
<td>traverse</td>
</tr>
<tr>
<td>alloy</td>
<td>consist</td>
<td>decrease</td>
<td>escort</td>
<td>increase</td>
<td>outgo</td>
<td>presage</td>
<td>remark</td>
<td>research</td>
<td>undress</td>
</tr>
<tr>
<td>ally</td>
<td>confuse</td>
<td>defect</td>
<td>essay</td>
<td>indent</td>
<td>outlaw</td>
<td>present</td>
<td>recount</td>
<td>reset</td>
<td>upcast</td>
</tr>
<tr>
<td>annex</td>
<td>conflict</td>
<td>defile</td>
<td>excerpt</td>
<td>inflex</td>
<td>outleap</td>
<td>produce</td>
<td>redraft</td>
<td>subject</td>
<td>uplift</td>
</tr>
<tr>
<td>assay</td>
<td>conscript</td>
<td>descant</td>
<td>excise</td>
<td>inflow</td>
<td>outlook</td>
<td>progress</td>
<td>refill</td>
<td>sublet</td>
<td>uprise</td>
</tr>
<tr>
<td>bombard</td>
<td>conscript</td>
<td>desert</td>
<td>exile</td>
<td>inlay</td>
<td>outpour</td>
<td>project</td>
<td>protest</td>
<td>refund</td>
<td>surcharge</td>
</tr>
<tr>
<td>cement</td>
<td>consort</td>
<td>detail</td>
<td>exploit</td>
<td>inlet</td>
<td>overspread</td>
<td>purport</td>
<td>protest</td>
<td>refuse</td>
<td>survey</td>
</tr>
<tr>
<td>collect</td>
<td>content</td>
<td>dictate</td>
<td>export</td>
<td>insert</td>
<td>outstretch</td>
<td>rampage</td>
<td>refuse</td>
<td>survey</td>
<td>suspect</td>
</tr>
<tr>
<td>combat</td>
<td>contest</td>
<td>digest</td>
<td>extract</td>
<td>inset</td>
<td>outwork</td>
<td>relief</td>
<td>remark</td>
<td>stomach</td>
<td>vomit</td>
</tr>
<tr>
<td>commune</td>
<td>contract</td>
<td>discard</td>
<td>ferment</td>
<td>insult</td>
<td>perfume</td>
<td>rebate</td>
<td>remark</td>
<td>stomach</td>
<td>vomit</td>
</tr>
<tr>
<td>contract</td>
<td>dispatch</td>
<td>double</td>
<td>handle</td>
<td>matter</td>
<td>present</td>
<td>remark</td>
<td>suspect</td>
<td>suspect</td>
<td>suspect</td>
</tr>
</tbody>
</table>

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

Script indicates pronunciation from (Boyer, 1700), *=1700, 1847 (James and Molé, 1847), and 2007 (Cambridge Advanced Learner’s Dictionary, OED) dictionaries are not identical.\(^{23}\)

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

\(^{23}\)Sample selection described in §2.