When does variation lead to change? A dynamical systems model of a stress shift in English

Morgan Sonderegger¹
(Department of Computer Science, University of Chicago)

Workshop on Language and Cognition
1/18/08

¹This work is in collaboration with Partha Niyogi
Outline

Introduction

Data
   Description
   Analysis

Models
   Model 1
   Model 2
   Interpretation

Conclusion
Variation and change

- All language change begins with variation between two (or more) forms, but not all variation leads to change.
- Write variation between old form $A$ and novel form $B$ as $A/B$, then can have:
  1. Variation disappears ($A/B \rightarrow A$)
  2. Variation persists ($A/B \rightarrow A/B$)
  3. Change occurs ($A/B \rightarrow B$)

- All outcomes can occur for similar variation patterns, or same one in different populations (e.g. Early Modern English: Nevalainen & Raumolin-Brunberg, 2003)

- **The big question:** Fine-grained variation occurs in every domain of language and language use, even within the speech of individual speakers (e.g. Pierrehumbert, 2003), yet does not usually lead to change. How do V & C coexist?
The "actuation problem", restated

1. Why does language change occur at all?
2. Why does it arise from variation?
3. What determines whether a pattern of variation is *stable* or *unstable* (leads to change)?
To explore these 3 questions, case study of stress shift in English, model its dynamics. Use 2 approaches to language change:

1. Building & analyzing diachronic data sets (sociolinguists: W.S.Y. Wang, W. Labov, ..).

Not previously combined, complementary strengths.

Theme: Long-term stability and sudden change coexist. Fits with dynamical system models containing bifurcations, corresponds to learners with "ambiguity".
Looking at English disyllabic, homographic noun/verb pairs (Lists 1, 2)

Productive class: rebound, party, YouTube.

But some go back to Old English:
ándsaca "apostate" vs. onsácan "to deny"

Three stress patterns observed:

<table>
<thead>
<tr>
<th>N</th>
<th>V</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>óσσ</td>
<td>elbow, fracture, forecast</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>óσσ</td>
<td>σσ</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>σσσ</td>
<td>σσσ</td>
</tr>
</tbody>
</table>

Conflict between Germanic, Romance stress rules.
Depending on method used to count, get $\approx 650 - 3000$ pairs.

Looking at a random subset of 100 of these words (List 2), most do *not* change stress over time.

But some do: Sherman (1975) counted 149 pairs (List 1) which have changed over 400 years.

Looked at dictionaries from 1570-1800, concluded that most changes were

1. $(1, 1) \rightarrow (1, 2)$
2. $(2, 2) \rightarrow (1, 2)$ (more frequent)

i.e. lexical diffusion to $(1, 2)$ ("diatone").

We filled in dataset to present day to understand the dynamics: diffusion, or more complicated?
Data collection

- Stress of the 149 words, N and V forms, examined in dictionaries.
- Sherman: every British dictionary with stress information, 1570-1799.
- New: 43 additional British and American dictionaries, 1800-2003, ~ every 5 years.
- 76 dictionaries total.
- Get 149 × 76 matrix of reported pronunciations (# words x # dictionaries), 63% full (words don’t exist yet, aren’t reported, etc., mostly earlier dictionaries).
Distribution of dictionaries

<table>
<thead>
<tr>
<th></th>
<th>1550-1699</th>
<th>1700-99</th>
<th>1800-99</th>
<th>1900-</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>British</td>
<td>8</td>
<td>25</td>
<td>15</td>
<td>14</td>
<td>62</td>
</tr>
<tr>
<td>American</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>10</td>
<td>14</td>
</tr>
</tbody>
</table>

- Two datasets (British and American), only British today.
- Full bibliographic information:
  people.cs.uchicago.edu/~morgan/diatones/dictionary_list.html
Methodological detour

How can we be sure dictionaries – especially older ones – are descriptive, not prescriptive?

Can’t, but encouraging evidence:

1. Prescriptivism is for regularity of pronunciation, but variation is recorded in all dictionaries examined.
2. No trend in (entries listing variation)/(total listed entries) over time ($r = -0.04$).
3. Pronunciation rarely socially marked (for this class), see both in present and past (lack of comments).
4. Attitude evidence: dictionary authors of 18th, early-mid 19th century (pre-OED) concerned with accurately recording "polite usage", supported by comments.
5. Cases of lexical variation, change recorded carefully, sometimes with regret, N/V pairs a primary example.
1. "... The first pronunciation .. [(2,2)] is adopted by [9 lexicographers] and .. [(1,2)] by [4 lexicographers]. As this [noun] was derived from the verb, it had formerly the accent of the verb: and that this accent was the most prevailing, appears from the majority of authorities in its favour. But the respectable authorities for the second pronunciation, and the pretence of distinguishing it from the verb, may very probably establish it, to the detriment of the sound of the language, without any advantage to its signification." [Walker (1802): "protest"]

2. "The accent [(2,2)] is proper, but in the mercantile world the verb is very commonly made to bear the same accent as the noun [(1,2)]." [Smart (1836): "discount"]

3. "Although all the orthoepists accent this word on the second syllable, yet we often hear it pronounced with the accent on the first." [Worcester (1859): "recess"]
Patterns of change

► Trajectories of moving average of pronunciation of each N/V pair constructed.

► In graphs (below), window for MA is 50 years, a point recorded at time $t$ if $\geq 2$ dictionaries have entries in $(t - 25, t + 25)$.

► Define "endpoints" as $(N(t), V(t)) = (1, 1), (1, 2), \text{or} (2, 2)$ (i.e. no variation).

► If "change"=move from one endpoint to another (conservative), 4 changes observed:
  1. $(1, 1) \rightarrow (1, 2)$
  2. $(1, 2) \rightarrow (1, 1)$
  3. $(2, 2) \rightarrow (1, 2)$
  4. $(1, 2) \rightarrow (2, 2)$
Change from $(1, 1)$ to $(1, 2)$

"combat"

Moving average of stress placement

Date

1700 1750 1800 1850 1900 1950 2000
Change from \((1, 2)\) to \((1, 1)\)
Change from \((2, 2)\) to \((1, 2)\)
Change from \((1, 2)\) to \((2, 2)\)
Other observations

- Change often, but not always to \((1, 2)\), more like an open migration process than diffusion:
Cycles possible: (noisy) examples observed.
Long-term variation (rare)

- Often see short-term variation from endpoints, rarely long-term (below).
Short-term variation (more common)
Main observations:

1. All change takes place through (1, 2) (not directly between (2, 2) and (1, 1)).
3. The pattern (2, 1) never occurs.
4. Stable, long-term variation almost never occurs, but lots of short-term variation around stable endpoints.

How can these patterns be explained, in particular 4.? Coexistence of stable states, variation around them, and rapid change between them, sometimes after 100s of years of stability.

How can a sudden loss of stability be explained? Models serve as testing ground for theories.
Modeling: What type of variation?

- What kind of variation: within or between individuals?
- Two forms, let $\alpha_i$ be the probability with with individual $i$ produces form 2. Possibilities:
  1. $\alpha_i \in \{0, 1\}$
  2. $\alpha_i \in [0, 1]$
  3. $\alpha_i \in \{0\} \cup (a, 1 - b) \cup \{1\}$

- Structure of variation has signif. consequences for model dynamics, today considering just $\alpha_i \in [0, 1]$, as suggested by following test.
Toy test: individual variation on NPR

- Find average for same speaker speaking same word. Similar for "perfume", "address" so far.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Word</th>
<th>N stress</th>
<th>V stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4</td>
<td>research</td>
<td>1.00 (25)</td>
<td>n/a</td>
</tr>
<tr>
<td>F1</td>
<td>research</td>
<td>1.00 (11)</td>
<td>n/a</td>
</tr>
<tr>
<td>F2</td>
<td>research</td>
<td>1.00 (6)</td>
<td>n/a</td>
</tr>
<tr>
<td>F3</td>
<td>research</td>
<td>1.00 (13)</td>
<td>1.00 (1)</td>
</tr>
<tr>
<td>F5</td>
<td>research</td>
<td>1.00 (10)</td>
<td>1.00 (3)</td>
</tr>
<tr>
<td>F6</td>
<td>research</td>
<td>1.00 (6)</td>
<td>1.00 (1)</td>
</tr>
<tr>
<td>M2</td>
<td>research</td>
<td>1.33 (6)</td>
<td>n/a</td>
</tr>
<tr>
<td>M3</td>
<td>research</td>
<td>1.44 (9)</td>
<td>n/a</td>
</tr>
<tr>
<td>M5</td>
<td>research</td>
<td>1.73 (11)</td>
<td>n/a</td>
</tr>
<tr>
<td>F4</td>
<td>research</td>
<td>1.90 (10)</td>
<td>n/a</td>
</tr>
<tr>
<td>M1</td>
<td>research</td>
<td>2.00 (7)</td>
<td>n/a</td>
</tr>
<tr>
<td>M6</td>
<td>research</td>
<td>2.00 (5)</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Model 1: psycholinguistic motivation

- Series of studies (Kelly 1988, 1989, Kelly & Bock 1988..):
- For real and novel words, nouns occur more often in trochaic-biasing than iambic-biasing contexts. Opposite true for verbs.
- Kelly et. al. showed this biases perception of nouns (trochaically) and verbs (iambically).
- Some stimuli:
  1. Trochaic bias, N: "Use the colvane proudly."
  2. Iambic bias, N: "The plants fontrain Joanne."
  3. Trochaic bias, V: "Gold will ponsect kingdoms."
  4. Iambic bias, V: "The dukes corvoot conceit."
- Implement this as "mishearing probability" that 2 heard given that 1 intended, etc.
Model 1: variables

- Infinite population size.
- Discretized generations: generation at $t + 1$ learns from generation at $t$.
- Consider 1 N/V pair. Each speaker keeps values $\tilde{\alpha}$, $\tilde{\beta} \in [0, 1]$ denoting how often produces the 2 form for nouns, verbs.
- At time $t$, let
  1. $\alpha_t$: probability a random noun example at $t$ is produced with final stress ($= 2$)
  2. $\beta_t$: same for verb ex.
Mishearing probabilities: let

\[
\begin{align*}
a_1 &= P(N \text{ heard as } 1 \mid 2 \text{ intended}) \\
b_1 &= P(N \text{ heard as } 2 \mid 1 \text{ int}) \\
a_2 &= P(V \text{ heard as } 1 \mid 2 \text{ int}) \\
b_2 &= P(V \text{ heard as } 2 \mid 1 \text{ int})
\end{align*}
\]

Then the probabilities a noun/verb example at \( t \) is heard as 2 are:

\[
\begin{align*}
P_1(t) &= \alpha_t (1 - a_1) + (1 - \alpha_t) b_1 \\
P_2(t) &= \beta_t (1 - a_2) + (1 - \beta_t) b_2
\end{align*}
\]
Model 1: learner

- Each learner hears $N_1$ noun examples, $N_2$ verb examples.
- Of these, $K_1$ nouns, $K_2$ verbs have final stress.
- $K_1 = K_1(t)$ and $K_2 = K_2(t)$ are random variables, each learner one sample.
- Batch learner: After hearing all examples, each learner sets
  \[
  \tilde{\alpha} = \frac{K_1}{N_1}, \quad \tilde{\beta} = \frac{K_2}{N_2}
  \]
- The expectation of the learners’ values gives $\alpha$ and $\beta$ for the next generation, i.e.:
  \[
  \alpha_{t+1} = E\left(\frac{K_1}{N_1}\right), \quad \beta_{t+1} = E\left(\frac{K_2}{N_2}\right)
  \]
- To take these expectations, have
  \[
  K_1 \sim \text{Bin}(P_1(t), N_1), \quad K_2 \sim \text{Bin}(P_2(t), N_2)
  \]
Model 1: Results

- Get iterated maps:
  \[
  \begin{align*}
  \alpha_{t+1} &= f_1(\alpha_t) := \alpha_t(1 - a_1) + (1 - \alpha_t)b_1 \\
  \beta_{t+1} &= f_2(\beta_t) := \beta_t(1 - a_2) + (1 - \beta_t)b_2
  \end{align*}
  \]

- Want fixed points: \( f_1(\alpha^*) = \alpha^* \), \( f_2(\beta^*) = \beta^* \):
  \[
  \alpha^* = \frac{b_1}{a_1 + b_1}, \quad \beta^* = \frac{b_2}{a_2 + b_2}
  \]

- Unique, stable fixed points, depend on \( a_i/b_i \) ratios... but this doesn’t explain sudden change.
- Similar for \( f_1, f_2 \) any linear combination of \( \alpha \) & \( \beta \).
Model 2

- Try another type of error: no mishearing, but an example can be heard as 1, 2, or *ambiguous*, in which case discarded.

- Consider just one form, same population assumptions as in Model 1, let $\alpha_t$ be the probability a random example produced as 2 at $t$.

- Ambiguity:

  $$ r_i = P(\text{heard as ambiguous} \mid i \text{ intended}) \quad (i = 1, 2) $$

- For a random example heard at $t$, let $P_i(t) = P(\text{heard as } i)$:

  $$ P_1 = (1 - \alpha)(1 - r_1), \quad P_2 = \alpha(1 - r_2) $$
Learner estimates $\tilde{\alpha}$

1. Hears $N$ examples: $K_1$ heard as 1, $K_2$ as 2, $N - K_1 - K_2$ ambiguous.
2. Sets

$$\tilde{\alpha} = \begin{cases} 
\frac{K_2}{K_1 + K_2} & \text{if } K_1 + K_2 > 0 \\
0 & \text{if } K_1 + K_2 = 0
\end{cases}$$

$z$ used if no unambiguous examples heard, can set to $\frac{1}{2}$.

For large $N$, can show that

$$E(\tilde{\alpha}) = \frac{E(K_2)}{E(K_1) + (K_2)}$$

$$\implies \alpha_{t+1} = f(\alpha_t)$$

$$:= \frac{\alpha(1 - r_2)}{(1 - r_1) + \alpha(r_1 - r_2)}$$
Get fixed points $x_\pm$:

\[ x_+ = 1 \quad \text{stable for } r_1 > r_2 \]
\[ x_- = 0 \quad \text{stable for } r_1 < r_2 \]

Bifurcation at $r_1 = r_2$, explains sudden change as loss of stability of a f.p.

This is simplest ambiguity model: by making more complicated, get more realistic behavior.

Can make $N$ finite, still get bifurcation-like behavior + frequency effect.

Mixture of ambiguity and mishearing: let

- $R$ be % of errors which are mishearing
- relative error=$(\text{mean error in hearing 1})/(\text{m.e. 1 + m.e. 2})$. 
Mixture model

- $R$ determines how "bifurcation-like" curve is.
Recap

- Ambiguity in model $\iff$ bifurcation, mechanism for sudden change in stability of a fixed point $=$ stability of variation.
- No ambiguity $\iff$ no sudden change.
- Can directly relate parameters to shape of modeled trajectories $\iff$ to trajectories for individual words.
- Interpretation of ambiguity?
To do..

- **Coupling**: Interaction between $N$ and $V$ dynamics not yet captured, absence of $(2, 1)$?

- **Morphology/Analogy**: Prefix classes important: seems that words with same prefix move together (trajectory distance) – weak/strong? Note almost all words in List 1 prefixed, many in List 2 not.

- **Frequency**: Word frequency often invoked w.r.t. lexical change, analogical vs. phonetic (e.g. Phillips 2006), role here?

- Effects of finite population size, non-overlapping generations, network structure...
But: Most of these issues need study more generally!
Study of language change incorporating both modeling, data still at early stage – hopefully have at least shown it’s a worthwhile direction, lots of potential for understanding structure of change.

Thanks!
References

List 1: 149 N/V pairs which have shown variation since 1570

<table>
<thead>
<tr>
<th>abstract</th>
<th>confine</th>
<th>discard</th>
<th>increase</th>
<th>outwork</th>
<th>recoil</th>
<th>surcharge</th>
</tr>
</thead>
<tbody>
<tr>
<td>accent</td>
<td>conflict</td>
<td>discharge</td>
<td>indent</td>
<td>perfume</td>
<td>record</td>
<td>survey</td>
</tr>
<tr>
<td>addict</td>
<td>conscript</td>
<td>discord</td>
<td>infix</td>
<td>permit</td>
<td>recount</td>
<td>suspect</td>
</tr>
<tr>
<td>address</td>
<td>conserve</td>
<td>discount</td>
<td>inflow</td>
<td>pervert</td>
<td>redraft</td>
<td>torment</td>
</tr>
<tr>
<td>affix</td>
<td>consort</td>
<td>discourse</td>
<td>inlet</td>
<td>post-date</td>
<td>redress</td>
<td>transfer</td>
</tr>
<tr>
<td>affect</td>
<td>content</td>
<td>egress</td>
<td>insert</td>
<td>prelude</td>
<td>refill</td>
<td>transplant</td>
</tr>
<tr>
<td>alloy</td>
<td>contest</td>
<td>eject</td>
<td>inset</td>
<td>premise</td>
<td>refit</td>
<td>transport</td>
</tr>
<tr>
<td>ally</td>
<td>contract</td>
<td>escort</td>
<td>invert</td>
<td>present</td>
<td>refund</td>
<td>traverse</td>
</tr>
<tr>
<td>annex</td>
<td>contrast</td>
<td>essay</td>
<td>invert</td>
<td>present</td>
<td>refusal</td>
<td>traverse</td>
</tr>
<tr>
<td>assay</td>
<td>converse</td>
<td>excerpt</td>
<td>misprint</td>
<td>present</td>
<td>refuse</td>
<td>traverse</td>
</tr>
<tr>
<td>bombard</td>
<td>convert</td>
<td>excise</td>
<td>legate</td>
<td>produce</td>
<td>regress</td>
<td>undress</td>
</tr>
<tr>
<td>cement</td>
<td>convict</td>
<td>exile</td>
<td>misprint</td>
<td>produce</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>collect</td>
<td>convoy</td>
<td>exploit</td>
<td>object</td>
<td>progress</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>combat</td>
<td>decoy</td>
<td>export</td>
<td>outcast</td>
<td>protest</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>commune</td>
<td>decrease</td>
<td>extract</td>
<td>outcry</td>
<td>purport</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>compact</td>
<td>defect</td>
<td>ferment</td>
<td>outlaw</td>
<td>rampage</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>compound</td>
<td>defile</td>
<td>impact</td>
<td>outleap</td>
<td>rebate</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>compress</td>
<td>descant</td>
<td>import</td>
<td>outlaw</td>
<td>rebel</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>concert</td>
<td>desert</td>
<td>impress</td>
<td>outpour</td>
<td>rebound</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>concrete</td>
<td>detail</td>
<td>incense</td>
<td>outpour</td>
<td>recall</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>conduct</td>
<td>dictate</td>
<td>incense</td>
<td>outpour</td>
<td>recast</td>
<td>upcast</td>
<td>upcast</td>
</tr>
<tr>
<td>confect</td>
<td>digest</td>
<td>incline</td>
<td>outstretch</td>
<td>recess</td>
<td>upcast</td>
<td>upcast</td>
</tr>
</tbody>
</table>
List 2: 100 randomly chosen N/V pairs from present-day English (*=no stress change since 1700)

abuse* ally anchor* arrest* attack* backpack* badger* bankrupt* beaver* bellow* blunder* buffer* cascade* centre* challenge* channel* discharge dissent* divide* elbow* entrance express* forecast fracture* fragment gallop* giggle* glimmer* glory* grumble* handle* highlight* import import index iron* levy* licence* matter* measure* merit* mirror* motion* motor* murder* notice* outline* paper* partner* party* patent* pattern* pencil* pervert* police premise proceed purchase* refund reject* relapse remark* repeal* repute* reserve* review* rival* safeguard* sandwich* scatter* second* signal* spiral* squabble* stable* swivel* throttle* travel* treble* triple* triumph* trouble* upset vomit* zigzag*