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

Partha Niyogi, Morgan Sonderegger
(Department of Computer Science, University of Chicago)

Symposium on Phonologization
4/26/08
Variation and change

Given variation between old form $A$ and novel form $B$ as $A/B$, can have:

1. Variation disappears ($A \sim B \rightarrow A$)
2. Variation persists ($A \sim B \rightarrow A/B$)
3. Change occurs ($A \sim B \rightarrow B$)

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

Q: Fine-grained phonetic variation extensive (e.g. Pierrehumbert, 2003), yet usually does not phonologize (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, ..).
2. Analytical modeling using dynamical systems (Hofbauer & Sigmund 1988 (population genetics), Niyogi & Berwick 1995, Niyogi 2006 (language)).

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".
Background: English noun-verb pairs

- Looking at English disyllabic, homographic noun/verb pairs (Lists 1, 2), productive class, Germ. vs. Rom. origin.
- Three stress patterns observed:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>V</th>
<th>(elbow, fracture, forecast)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>ơơ</td>
<td>ơơ</td>
<td>(consort, protest, refuse)</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>ơơ</td>
<td>ơơ</td>
<td>(cement, police, review)</td>
</tr>
</tbody>
</table>

- \( \exists \approx 650 - 3000 \) pairs.
- In random subset (List 2), most do not change over time.
- Some do: Sherman (1975) counted 149 pairs (List 1) changed over 400 years.
- Dictionaries 1570-1800, concluded that most changes \((1, 1) \rightarrow (1, 2)\) or \((2, 2) \rightarrow (1, 2)\) (more freq), i.e. lexical diffusion to \((1, 2)\)
- 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, 8 1550-1699
- Get $149 \times 76$ matrix of reported pronunciations (# words x # dictionaries), 63% full (words don’t exist yet, aren’t reported, etc., mostly earlier dictionaries).
- Only British dictionaries used today.
- Qualitative and quantitative methodology checks for prescriptive vs. descriptive.
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 $(N(t), V(t)) = (1, 1), (1, 2), (2, 2)$ (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 $(1, 2)$
Change from \((1, 2)\) to \((1, 1)\)
Change from \((2, 2)\) to \((1, 2)\)
Change from (1, 2) to (2, 2)
Diffusion?

- Change often, not exclusively to \((1, 2) \implies \) more complicated model needed (open migration process?)
Cycles possible: (noisy) examples observed.
Short-term variation in N or V around endpoints (more common)
Long-term variation (rare)
Main observations:

1. All change takes place through (1, 2) (not (2, 2) ↔ (1, 1)).
2. “Multidirectional”?
3. The pattern (2, 1) never occurs.
4. Stable, long-term variation almost never occurs, but lots of short-term variation around stable endpoints.

- Coexistence of stable states, variation around them, and rapid change between them, sometimes after 100s of years of stability.
- How to explain sudden loss of stability? Modeling 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 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.
Toy test: individual variation on NPR

- Same speaker using N “research”, similar pattern for "perfume", "address", “increase”.

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Dialect</th>
<th>N=1</th>
<th>N=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>ira</td>
<td>A</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>joanne</td>
<td>A</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>margery</td>
<td>A</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>kathyc</td>
<td>A</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>michelen</td>
<td>A</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>jonathanm</td>
<td>A</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>bbc1</td>
<td>B</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>tim</td>
<td>B</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>sprakash</td>
<td>I</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>davidk</td>
<td>A</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>anthonyr</td>
<td>A</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>eric</td>
<td>A</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>joe</td>
<td>A</td>
<td>4</td>
<td>9</td>
</tr>
<tr>
<td>ukprof</td>
<td>B</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>marys</td>
<td>A</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>dang</td>
<td>A</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td>paulw</td>
<td>A</td>
<td>0</td>
<td>5</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).
- 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

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

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

\[ P_1(t) = \alpha_t (1 - a_1) + (1 - \alpha_t) b_1 \]
\[ P_2(t) = \beta_t (1 - a_2) + (1 - \beta_t) b_2 \]
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:

\[
\alpha_{t+1} = f_1(\alpha_t) := \alpha_t(1 - a_1) + (1 - \alpha_t)b_1 \\
\beta_{t+1} = f_2(\beta_t) := \beta_t(1 - a_2) + (1 - \beta_t)b_2
\]

- 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 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 $P_i(t) = P(\text{heard as } i)$:

$$P_1 = (1 - \alpha)(1 - r_1), \quad P_2 = \alpha(1 - r_2)$$
Model 2: modified batch learner

▶ 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

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

$z$ 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 = (mean error in hearing 1)/(m.e. 1 + m.e. 2).
Mixture model

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

- Ambiguity in model $\implies$ bifurcation, mechanism for sudden change in stability of a fixed point = stability of variation.
- No ambiguity $\implies$ no sudden change.
- Can directly relate parameters to shape of modeled trajectories $\implies$ to trajectories for individual words.
- Interpretation of ambiguity?
Not discussed/To do

- **Coupling**: Interaction between $N$ and $V$ dynamics, absence of $(2, 1)$? Have models, hard to get both no $(2, 1)$ and bifurcations w/o hardwiring.

- **Morphology/Analogy**: Prefix classes important: using a distance metric on trajectories, get graph with nodes=words, can see how much words sharing prefix have similar trajectories, find that words in large prefix classes have more similar trajectories ($R = 0.69$, $p < 0.05$) Note almost all words in List 1 prefixed, many in List 2 not.

- **Frequency**: Word frequency in lexical change, analogical vs. phonetic (e.g. Phillips 2006), role here? May be trigger for bifurcation, along with $K$. To do.

- **Effects of finite population size, non-overlapping generations, network structure...**
 Conclusion

- All these issues need study more generally.
- Study of change incorporating both modeling and data still at early stage – hope have shown it’s worthwhile, lots of potential for understanding structure of change using mathematical/computational tools, esp. actuation (often unobserved), phonologization/grammaticization, transmission.
- Thanks!
References

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

abstract  confine  discard  increase  outwork  recoil  surcharge
accent   conflict discharge indent perfume record  survey
addict   conscript discord infix permit recount  suspect
address  conserve discount inflow pervert redraft  torment
affix    consort discourse inlay post-date redress  transfer
affect   content egress insert prelude refill  transplant
alloy    contest eject inset premise refund  transport
ally     contract escort insert present refit  traverse
annex    contrast essay invert present refuse  traverse
assay    converse excerpt invert presume regress  undress
bombard  convert excise legate produce rehash  upcast
cement  convict exile misprint protest reject  upgrade
collect  convoy exploit object relapse relay  uplift
combat   decoy export object repeat rise  uprise
commune decrease extract outcast relapse reprint  uprush
compact  defect ferment outcry rebate sojourn  subject
compound defile impact outlaw rebel reset  sub-let
compress descant import outleap rebound sojourn  sublease
concert  desert impress outlook recall research  subject
concrete detail imprint outpour recast reprint  sub-let
conduct  dictate incense outspread recast reprint  sub-let
confect  digest incline outstretch recess  sub-let
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 index iron* levy* licence* measure* merit* mirror* motion* motor* murder* notice* outline* paper* partner* party* patent* pattern* pencil* pervert* police premise prune* 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*