Predicting Language Change

Charles Yang
Department of Linguistics & Computer Science
Institute for Research in Cognitive Science
University of Pennsylvania

PhyloLing, Ragusa
Language & Evolution

• Darwin: “A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue.” (Descent)

• Forward and backward approaches to change:
  • population genetics & phylogenetic methods
Mechanisms of transmission may be different for language & genes (cf. Cavalli-Sforza & Feldman 1981)

Development of a population genetic theory of language change
Natural Selection

• Consider two alleles (red vs. green eyes) A and a, which take up proportion of p and q in a population

• Suppose A is a little better than a; the former has a fitness of 1 and the latter (1-s): in nature, s is usually very small

• Fitness values can be measured either naturally or experimentally

• Basic models of population genetics were developed well before the biochemical articulation of genetic structures (or “parameters”)

Basic Model of Selection

- Current generation
  - p and q
- Fitness and Selection (1: 1-s)
  - A: p
  - a: q(1-s)
- Next generation

\[
p' = \frac{p}{p + q(1 - s)}
\]
The Variational Model

- Two “grammars” A and a, with probability p and q
  - innate or learned
- Upon hearing an input token, select A/a with prob. p/q
  (Bush & Mosteller 1951): *Reinforcement Learning*

\[
\begin{align*}
\text{if } A \rightarrow s \text{ then} & \quad \begin{cases} 
    p' = p + \gamma q \\
    q' = (1 - \gamma)q
\end{cases} \\
\text{if } A \not\rightarrow s \text{ then} & \quad \begin{cases} 
    p' = (1 - \gamma)p \\
    q' = q + \gamma p
\end{cases}
\end{align*}
\]
Grammar and Cheese

Bush & Mosteller (1951)
Fitness

• **Penalty probability**: the prob. with which a grammar clashes with input

  • something we can obtain from corpora, not for the learner to track off

• **Convergence**: let $C_A$ and $C_a$ be the penalty probability

  • Interested in the predictable outcome of competing variants

\[
p = \frac{C_a}{C_A + C_a} \quad q = \frac{C_A}{C_A + C_a}
\]
Basic model of change

$$C_A = \beta q, \quad C_a = \alpha p$$

$$p' = \frac{\alpha p}{\alpha p + \beta q}$$

$$p' = \frac{p}{p + (1 - s)q}$$

where

$$s = \frac{\alpha - \beta}{\alpha}$$
Language change is natural selection

- If acquisition is variational, then the S-shape must be hold for language change
Two Kinds of Changes

- Morphological and syntactic change (Kroch 1989, etc.)
  \[ (p_0, q_0) \rightarrow (p_1, q_1) \rightarrow \ldots \rightarrow (p_n, q_n) \]

- Phonemic changes tend to be abrupt (Herold 1990, Johnson 2007; see also Lightfoot 1991), perhaps reflecting the limit of bilingualism in the phonetic system (Cutler et al. 1989)
  \[ (p_0, q_0) \rightarrow p_1 \text{ if } p_1 > q_1: \text{ winner take all} \]

- Two cases of language change where the outcome (and time course) are predictable
Case I: Old to Modern French

- French went from a V2 Grammar to a -V2 grammar and lost pro-drop as well (“parameters”)

Loss of null subjects

a. *Ainsi s’amusaient bien cette nuit. (ModF)
   thus (they) had fun that night
b. Si firent grant joie la nuit. (OF)
   thus (they) made great joy the night

Loss of V2

a. *Puis entendirect-ils un coup de tonnerre. (ModF)
   then heard-they a clap of thunder
b. Lors oïrent ils venir un escoiz de tonoire. (OF)
   then heard they come a clap of thunder
The competition

• V2 went against SVO and lost

• But corpus studies of V2 grammars (Dutch, German) and SVO grammars (English) show the contrafactual results:
  • $\sim 10\%$ of SVO grammar sentences ($SXVO$ and $XSVO$) are against V2: $\alpha = 0.1$
  • 70% of V2 grammar sentences are SVO and the rest are against SVO: $\beta = 0.3$
  • $\alpha < \beta$: French could not have evolved from Old to Modern
Pro-drop: The culprit

Quickly read the book (Good for SVO)

read the book (Bad for V2)

Actual competition between (pd, V2) and (pd, SVO)
\[ \alpha' = 0.9n + 0.1(1 - n), \quad \beta' = 0.3(1 - n) \]
Direction of Change

- Change happens if $\alpha' > \beta'$, or $n > 2/11$ (0.182)
  - If the frequency of null subjects ($n$) during the history of French exceeds 18% of all subjects, then **V2 must** lose to **SVO**.

- This condition was met in 14th century French (data from Roberts 1993)

<table>
<thead>
<tr>
<th>Text</th>
<th>SV</th>
<th>VS</th>
<th>NullS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Froissart, <em>Chroniques</em> (c. 1390)</td>
<td>40%</td>
<td>18%</td>
<td>42%</td>
</tr>
<tr>
<td><em>15 Joyes (14esme Joye)</em> (c. 1400)</td>
<td>52.5%</td>
<td>5%</td>
<td>42.5%</td>
</tr>
<tr>
<td>Chartier <em>Quadrilogue</em> (1422)</td>
<td>51%</td>
<td>7%</td>
<td>42%</td>
</tr>
</tbody>
</table>

- All western Romance languages that had V2 order lost it: all had pro-drop (like French) and many still have (Spanish, Italian, Catalan ...)

Time course of change

\[ p' = \frac{p}{p + (1 - s)q} \]

\[ q' = \frac{(1 - s)q}{p + (1 - s)q} \]

\[ \frac{q'}{p'} = (1 - s) \frac{q}{p} \]

\[ \frac{q_t}{p_t} = (1 - s)^t \frac{q_0}{p_0} \]

\[ t = \frac{\log \frac{q_t}{p_t} - \log \frac{q_0}{p_0}}{\log(1 - s)} \]

- Let beginning of change \((p_0)\) be 1%, and end \((p_t)\) be 99%
- The number of generation to completion \(t = -4/\log(1-s)\)
Time Course of Change

\[ \alpha' = 0.9n + 0.1(1 - n), \quad \beta' = 0.3(1 - n) \]

- % null subjects = 0.42 = \(n\) (Roberts 1993)
- \(s = (\alpha' - \beta')/\alpha' = 0.6, \ t = -4/\log(0.4)\)
- \(t \approx 10\) generations or 200-300 years for V2 to be lost, which seems largely consistent with the historical record
  - (\(n\) itself may be changing)
- Crucially pro-drop was lost much later in French

<table>
<thead>
<tr>
<th>Time Period</th>
<th>SV</th>
<th>VS</th>
<th>pro</th>
</tr>
</thead>
<tbody>
<tr>
<td>15th century</td>
<td>48%</td>
<td>10%</td>
<td>42%</td>
</tr>
<tr>
<td>16th century</td>
<td>77%</td>
<td>3%</td>
<td>15%</td>
</tr>
</tbody>
</table>
Cot vs. Caught


Aaron Dinkin (Beverly, MA): I pronounce *caught* the same as *cot*
When *cot* Met *caught* (in one family!)

- Tom: 43
- Lonnie: 42
- Amber: 16
- Sharon: 9

Seekonk MA

Courtesy of Daniel Ezra Johnson
• Pose challenges to the view that language functions/changes to facilitate communication (Labov 1994, 2001)

• Mergers, and phonemic systems in general, are acquired relatively early (Payne 1976) and not generally susceptible to socio-indexical valuation

• Mergers are among the best documented linguistic changes on record

• Mergers are rarely, if ever, reversed (Labov 1994)
Bad Vowels

• Miscommunication? Actually very rare ...

• Labov (2001, 2010): just dozens of examples
  
  • $V_1$: It would be even better if Don could take her to the airport
  
  • $V_2$: (wondering how Dawn, who is blind, could take anyone to the airport)

• $V_2$ listeners have far greater error rates than $V_1$ listeners ($\approx 4:1$)
Delay as Penalty

- In speech/language processing, the phonological form of words is accessed first, prior to the integration of other structural and contextual cues.

- For homophonous words, frequency effects determine the time course of semantic activation.

- Boland & Blodgett (2001, *J. Mem. Lg.*), among many others
  - She saw him *duck* and stumble near the barn.
  - Significant delay in processing even when the contexts are clear: \( \text{duck}_N \gg \text{duck}_V \)
  - “I slept on a cot”: A Bostonian first hears *caught* (which sounds the same)
Penalizing V1

- You only hear one vowel (both o and aw are aw)
- Assume Don is more frequent than Dawn

\[ C_1 = \sum \min(f_o, f_{aw}) \]
Penalizing V2

- you have two vowels (o and aw) but they are psycho-acoustically confusable

<table>
<thead>
<tr>
<th></th>
<th>o</th>
<th>aw</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>92.3</td>
<td>3.5</td>
</tr>
<tr>
<td>aw</td>
<td>13.8</td>
<td>82.0</td>
</tr>
</tbody>
</table>

Hillenbrand et al. (1995, JASA)

\[ C_2: \text{linear function of frequency, confusability, and } P_0 \]
• Frequency counts from various corpora consistently produce threshold in 17-23% range
From the field

- Johnson (2007: p248) describes 3 stages
  1. “... not many merged parents in the community ... small number of merged children will learn the distinction from their peers”
  2. “more merged parents have moved in ... The proportion of natively merged children entering the peer group ... not enough to stop natively distinct children remaining distinct”
  3. “the proportion of natively merged children exceed Y [CY: a threshold]. While distinct children may not be in a minority, they have enough contact with merged peers that they lose their inherited distinction”
- School Survey: 18% of merged peers for 12th grade, 20% for 8th, and 23% for 4/5th.
Beyond the field

• Accounts for the easy spreadability of mergers (Herold 1990)
• And the irreversibility of mergers: you need, e.g., 4 times of the migration population to the native one

• Accounts for the resistance to mergers in in urban areas with large populations (mid-Atlantic: NJ, Philadelphia, DC)

• Effect of mutual confusability: important but not critical
  • confusability = 0.01 both ways, $p_0 = 40\%$
  • no confusability, $p_0 = 43\%$
  • mergers are inherently favored because of the nature of language processing: phonology first, context later and frequency rules
**Pin-Pen merger**

- minimal pairs: din-den, tin-ten, gin-Jen/Gen, pin-pen, jim-gem, bin-ben ... 
- different words & confusion matrix

/i/ and /ɛ/ before nasals

\[ p_0 \approx 32\% \]  
merger will spread!
Pin-Pen merger

- Brown (1991: *American Speech*): spread was abrupt
- Source of data: dialect surveys from TN and NC
/l-n/ merger in HK Cantonese

• Merger originated in Guangdong Province and spread to HK via migration (Zee 1999, Bourgerie 1990)

• Different words, different confusion matrix (Wang 2007)

• Tipping point: 27% (merger is sudden)

• Roughly a quarter of the 31-45 years’ parents were from merged areas

<table>
<thead>
<tr>
<th>Age</th>
<th>Number</th>
<th>Mean frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-18</td>
<td>12</td>
<td>0.854</td>
</tr>
<tr>
<td>19-30</td>
<td>22</td>
<td>0.700</td>
</tr>
<tr>
<td>31-45</td>
<td>36</td>
<td>0.669</td>
</tr>
<tr>
<td>46+</td>
<td>11</td>
<td>0.282</td>
</tr>
</tbody>
</table>

from Bourgerie (1990)
Conclusion

• Projecting the present to the past
  • A synthetic approach to change with theories of language structure, processing and transmission
  • Complements the phylogenetic approaches to language change
• Predicting the future, predicting the past