Approaching Actuation

Charles Yang
University of Pennsylvania

University of York, Feb 2015
Actuation

• Weinreich, Labov & Herzog (1968)

• Origin of linguistic variation:
  • “they are there all along”: e.g., the biological space of grammar (WLH’s constraints)
  • Arise from linguistic transmission (i.e. experience dependent learning and generalization)
“Add -d” is not an innate UG option!
How to Generalize

- Let $R$ be a rule applicable to $N$ lexical items out of which $e$ do not follow $R$. $R$ is productive iff

$$e = \frac{N}{\ln N}$$

- The number of tolerable exceptions must be relatively small: otherwise the learner must lexicalize everything

- English past tense: $e = 120$, $N$ must be greater than 800 for -ed to be productive

- Emergence of “-d” as a productive rule takes place only after the child’s regulars overwhelm irregulars
When Exceptions Overwhelm

- Polish **singular** masculine genitives take either -a or -u as suffix but neither seems to be the default based on a suite of tests (Dabrowska 2000).

- **Plurals** take -ow as the default, with exceptional -i/y suffix

- Analysis of child-directed Polish in CHILDES

<table>
<thead>
<tr>
<th>suffix</th>
<th>type freq.</th>
<th>productive?</th>
<th>ave. token freq.</th>
<th>error %</th>
</tr>
</thead>
<tbody>
<tr>
<td>-a (sg.)</td>
<td>837</td>
<td>no</td>
<td>7.17</td>
<td>1.28%</td>
</tr>
<tr>
<td>-u (sg.)</td>
<td>516</td>
<td>no</td>
<td>8.8</td>
<td>0.24%</td>
</tr>
<tr>
<td>-ow (pl.)</td>
<td>551</td>
<td>yes</td>
<td>6.5</td>
<td>0.41%</td>
</tr>
<tr>
<td>-i/y (pl.)</td>
<td>61</td>
<td>no</td>
<td>11.4</td>
<td>15.53%</td>
</tr>
</tbody>
</table>
Parameters and Exceptions

- Newmeyer (2004): parameters are wrong because they have exceptions
- A 200,000 word French directed corpus; adjectives tagged and lemmatized
- 20 exceptional adjectives
- 120 “regular” adjectives: 80 would suffice
- The default parameter value (or rule) can be maintained and the rest are lexicalized

une *vieille* amie ‘a friend for a long time’
une amie *vielle* ‘a friend who is aged’
Two case studies

- The learner creates systematic regularities which lead to change

- **PERmit** (noun) vs. **perMIT** (verb): diatonic stress

- Dative alternations
  - John donated a painting to the museum
  - *John donated the museum a painting*
Diatonic Stress

<table>
<thead>
<tr>
<th>N-V pairs</th>
<th>Total</th>
<th>Isotonic</th>
<th>Diatonic</th>
<th>% of Diatones</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disyllabic</td>
<td>1315</td>
<td>1165</td>
<td>150</td>
<td>11</td>
</tr>
<tr>
<td>Polysyllabic</td>
<td>442</td>
<td>372</td>
<td>70</td>
<td>16</td>
</tr>
<tr>
<td>Total</td>
<td>1757</td>
<td>1537</td>
<td>229</td>
<td>13</td>
</tr>
</tbody>
</table>

Sherman (1975)
English Stress

- The majority (>80%) of words are stress initial (Cutler & Davis 1987)
- Over 90% of bisyllabic nouns are trochaic, and about 70% of bisyllabic verbs are iambic (Bock & Kelly 1988)
- But English stress system is not a simple rule with 10% of exceptions
Bisyllabic Nonce Words

The second factor was Lexical Class. The two levels were (1) words produced in the noun sentence frame and (2) words produced in the verb sentence frame. In two syllable words, nouns are more likely to have penultimate stress than verbs (Kelly & Bock, 1988). If statistical generalizations about the distribution of main stress placement among lexical items have an effect on the placement of stress in novel words, then this factor should have a significant effect and nouns should receive more first syllable stress than verbs independent of the syllabic structure.

First consider the mean proportion of first syllable stress for the four syllable types and two lexical classes displayed in Figure 1. Note that, in general, nouns were more often produced with first syllable stress than verbs. Also note that a long vowel in the first syllable conditioned more first syllable stress, whereas a long vowel in the final syllable conditioned more final stress.

The main effects of Syllabic Structure and Lexical Class were both significant, $F(3, 72) = 53.39, p < .001$ and $F(1, 72) = 203.87, p < .001$ respectively, as was the interaction, $F(3, 72) = 27.06, p < .001$. The interaction was explored by investigating the effect of Syllabic Structure on noun and verb productions separately and by investigating the effect of Lexical Class on each of the four word types.

The effect of Syllabic Structure was significant for nouns, $F(3, 36) = 58.62, p < .001$. Tukey's tests ($p < .01$) revealed that words of the structure CVVCVCC (Type 1) were more...
Previous Work

- A paradigm case for *lexical diffusion* (Wang 1969 etc.)
  - Language change proceeds on a word by word basis, rather than through a systematic grammar
  - Heavily influenced the phylogenetic approach to language classification
- The role of frequency: Words (N/V pairs) with lower frequencies tend to change first (develop diatonic stress first; Phillips 1984, Bybee 2002, Lieberman et al. 2007)
Theories of English Stress

• Halle (1998): $\sigma_1 \sigma_2$

• Nouns
  • If $\sigma_1$ has VV or VC+ then it’s stressed (kingdom, beetle)
  • Else if $\sigma_2$ has VV then it’s stressed (device)
  • Else $\sigma_1$ is stressed (atom)

• Verbs
  • If $\sigma_2$ has VV or VCC+ then it’s stressed (revise, eject)
  • Else $\sigma_1$ is stressed (ambush)
14% vs. 6%

- Legate & Yang (2012: Tribute volume for Carol Chomsky)
- Words that appear at least once per million in child directed English (excluding proper names): 5763
- Initial stress: 4960 (14% exceptions): above threshold
- Halle (1998): 5408 (6% exceptions): below threshold
- Only the Halle (1998) account can productively tolerate the amount of exceptions
Hypothesis

• Diatonic stress shifts for words that are predicted to have divergence N/V stress

• (1, 2): construct
  • (1, 1): ambush
  • (2, 2): amount
  • (1, 2): reject
  • (1, 1): whisper

• (2, 1): 👻 (Sherman 1975, Sonderegger & Niyogi 2013)
Data

• CELEX homographic N/V pairs
• CMU Pronunciation Dictionary to obtain diatonic pairs
• SUBTLEXus corpus for word frequencies
• Halle (1998) predicts whether a specific N/V should shift or not

• 781 multisyllabic N/V pairs, 90 are diatonic (11.5%)
• 668 bisyllabic N/V pairs, 85 are diatonic (12.7%)
## Predicting Diatones

<table>
<thead>
<tr>
<th>Attested</th>
<th>Predicted Diatonic</th>
<th>Predicted Isotonic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diatonic</td>
<td>62</td>
<td>23</td>
</tr>
<tr>
<td>Isotonic</td>
<td>140</td>
<td>443</td>
</tr>
</tbody>
</table>

Fisher’s exact test: $p < 0.001$, odds ratio $= 8.50$
### Lexical Frequency?

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef.</th>
<th>S.E.</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-3.470</td>
<td>0.688</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>log(Freq)</td>
<td>0.089</td>
<td>0.111</td>
<td>0.422</td>
</tr>
<tr>
<td>Halle (1998)</td>
<td>3.159</td>
<td>0.855</td>
<td>&lt;0.001 ***</td>
</tr>
<tr>
<td>log(Freq) x Halle 1998</td>
<td>-0.179</td>
<td>0.141</td>
<td>0.204</td>
</tr>
</tbody>
</table>

![Graph showing frequency distribution](image_url)

- **Isotonic**
- **Diatonic**
*(2, 1)*

- No N/V pair has ever surfaced as (2,1) but there are still plenty of 2 nouns (today) and 1 verbs (follow)

- If a word does not follow the structural description of a productive rule, then no amount of lexical similarity can trigger change

- This seems like a radical statement for proponents of analogy but the child language evidence for lexically based analogy is virtually non-existent (Berko 1958, Marcus et al. 1992, Xu & Pinker 1995, Clahsen et al. 2002, Yang 2002, Guasti 2004 …)
Baker’s Paradox

a. John threw the ball to Bill.
   John threw Bill the ball.

b. John told Bill the story.
   John told the story to Bill.

c. John pushed the tea to Mary.
   *John pushed Mary the tea.

d. John donated the painting to the museum.
   *John donated the museum the painting.

• How to generalize just enough? (I texted him my apology)
Extension

- I pilked the cup to Petey ⟹ I picked Petey the cup
- I gorped Toby the key ⟹ I gorped the key to Toby

Conwell & Demuth (2007, Cognition)
Dative Acquisition

- Pinker (1989): weakly conservative learning
- Broad Range Rules: semantic properties necessary for dative constructions
- Narrow Range Rules: subclasses with different productivity
  - “future giving”: offer, promise, bequeath, leave, refer, forward, allocate, guarantee, award, reserve …
  - “continuous causation of accompanied motion”: pull, carry, push, schelp, lift, lower, haul …
  - “telecommunication”: telegraph, fax, telephone, email, wire …
The subjects in Conwell and Demuth (2007)'s extension study are on average only 3 years old, yet they generalize novel verbs in both constructions. For instance, when presented with "I pilked the cup to Petey", children can generalize "pilk" to the DOC expression "I pilked Petey the cup". Conwell and Demuth also notice a strong preference to extend a verb from DOC to PAC than from PAC to DOC, despite the fact that PACs are less frequent than DOCs in the input (Campbell and Tomasello 2001, Snyder and Stromswold 1997). Overall, children extended a novel verb from DOC to PAC 48.0% of the time but from PAC to DOC only 8.9% of time (though still significantly more than zero). (We provide an account for this finding in section 3.2 below.) Taken together, these studies suggest that dative constructions are productive in child language: children are not lexically conservative and they must confront Baker's Paradox head on (Pinker 1989).

3.2 Why do children “say me no”?

The allocated space does not allow for an extensive review of the literature on dative constructions and their acquisition. Rather, I will highlight some key points including Pinker's approach 1989 which has been influential in subsequent L1 and L2 research. There is now broad agreement on the semantic conditioning of these constructions (Goldberg 1995, Krifka 1999, Levin 1993, Hale and Keyser 2002, Beck and Johnson 2004). DOC verbs generally involve caused possession of the theme by the goal, understood in the broad including metaphorical sense. By contrast, PAC verbs tend to entail caused motion of the theme along the path to goal. But it is important to note that these are no more than necessary conditions on the availability of dative constructions, or what Pinker calls “broad-range rules”. There are language specific constraints that further restrict the availability of the datives. To this end, Gropen et al. (1989) propose a set of “narrow-range rules” that form the sufficient conditions. The contrast between (6a) and (6c) is accounted for because the former belongs to the class that describes “instantaneous causation of ballistic motion” (ibid., p243), which allows double object construction whereas the latter falls in the class of “continuous causation of accompanied motion in some manner” (ibid., p244) which does not.

The broad and narrow range rules are not a complete solution to the dative acquisition problem. First, while the broad range rule as a sufficient condition for datives seems well supported cross-linguistically, the distributions of the narrow range rules vary considerably cross languages (Jelinek and Carnie 2003, Jung and Miyagawa 2004) and must be acquired on the basis of the input data—especially since they include classes such as “instruments of communication” (radio, netmail, fax etc.) which cannot be directly attributed to Universal Grammar. Table 3 summarizes the cross-linguistic patterns of DOC (and its equivalent) for some of Gropen et al. classes (from Levin 2008; see references within):

<table>
<thead>
<tr>
<th>Class</th>
<th>Greek</th>
<th>English</th>
<th>Warlpiri</th>
<th>Hebrew</th>
<th>Icelandic</th>
<th>Mandarin</th>
<th>Yaqui</th>
<th>Fongbe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Give-type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Future having</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
</tr>
<tr>
<td>Send-type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>ND</td>
</tr>
<tr>
<td>Bring/take</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Y/N</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>ND</td>
</tr>
<tr>
<td>Throw-type</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Y/N</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Push-type</td>
<td>No?</td>
<td>No?</td>
<td>No</td>
<td>No?</td>
<td>ND</td>
<td>No</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

Datives must be acquired inductively on specific linguistic data

Children do not learn datives instantly (Snyder & Stromsworld 1995)
Double Object and P Construction

- Two million words of child directed English
- Extracted all “dativizable” verbs from the list from Gropen et al. (1989) and Levin (1993)
- Manually documented whether a verb was used in Double Object and/or Prepositional Construction Construction at least once
50 DOC-able verbs

- give, pass, hand, sell, pay, trade, lend, loan, serve, feed
- throw, toss, flip, kick
- send, mail, ship
- bring, take
- offer, promise, leave, refer, forward, allocate, guarantee, award
- tell, show, ask, teach, write, read, quote
- bake, make, build, cook, sew, knit, toss, fix, pour
- get, buy, find, steal, order, win, earn, grab
Productivity of DOC

- $N = 50 \ (e=13)$: 40 participate in DOC, 10 have not
  - DOC is a productive process in English
- $DOC \implies PC$: N=42, 40 participate in PC, 2 do not
  - Every DOC verb can be extended
*donate the museum the painting

- Latinate verbs generally don’t participate in DOC: children show sensitivity to these morphological/phonological properties in novel verb extension tasks (Gropen et al. 1989)

- Only a tendency:
  - Latinate: *assign, advance, award, guarantee, reserve* ...
  - Germanic: *shout, trust, lift, pick*, ...

- Levin (1993): 70 Latinate verbs, only 10 participate in double object construction (cannot be studied with CDS)

  - Non-productivity is productive

  - *assign, advance, award, guarantee, reserve* are positive counterexamples that we lexicalize
Approaching Actuation

- We now have reasonable descriptions, possibly even theories, of how children learn languages.
- Linguistic data for acquisition is finite and subject to sampling effects.
- Children will arrive at a range of productive hypotheses (e.g., $e < N/\ln N$), some of which will be ironed out, some will remain, including those that differ from the previous generation.
- Examination of the statistical composition of historical data may give us clues of variants that got actuated.