Learning Diacritics

Julie Anne Legate & Charles Yang
University of Pennsylvania

Parallel Domains
May 6th 2011
Where grammars leak

• “Clearly, we must design our linguistic theory in such a way that the existence of exceptions does not prevent the systematic formulation of those regularities that remain.” (SPE, p172)

  “basic word order”, “default rule”, “unmarked form”
  vs.
  “lexical listing”, “exceptional marking”, “diacritics”

• Need to find where grammars leak

  • General principles of language learning (aka Evaluation Procedure) that may apply across linguistic domains
Child Language: No gradience

- Rarely, if ever, confuse rules with exceptions
- Past tense acquisition
  - “bring-bringed” type: 10% errors (Marcus et al. 1992, Yang 2002)
  - “bring-brang” type: 0.2% errors (Xu & Pinker 1995)
  - Berko (1958): “binged” abundant, only one “brang” out of 84
- Raising of the main verb in English syntax is non-existent despite
  - “Baa Baa Blacksheep have you __ any wool?”
- English as a non-subject drop language despite
  - “seems nice to me”
- The acquisition of stress
Why Stress?

- Important disagreements remain
- Strong (longitudinal) studies of stress acquisition (e.g. Fikkert 1994, etc.)
- How does the child learn a mixture of both?
- How does the acquisition data inform the theory?
- How does the solution generalization across domains, from phonology to morphology to syntax?
Learning rules

• A general model for detecting productivity (Yang 2005)

• more exceptions ⇒ slower real-time processing: evidence from syntax (Swinney & Cutler 1979) and morphology (Sonnenstuhl & Huth 2002, Clahsen et al. 2002)

• Tolerance Principle: If a rule/process/morpheme/constraint could apply to N items, then the number of exceptions m it can tolerate while remaining productive is no more than

\[ \frac{N}{\ln N} \]

• For instance, if there are 120 irregular verbs in English, there needs to be 680 regular verbs to have a productive -ed rule: \( \frac{800}{\ln(800)} \approx 120 \)

• See Gorman & Yang (forthcoming) for additional cases including paradigmatic gaps (Halle 1973)
Tolerance for Birther

- **-er** is a productive suffix of English

- in morphological processing, we seem to strip off **-er** even in words such as “brother” leading to a garden path (Taft & Forster 1975, Rastle et al. 2004)

- But some **-er** are real/compositional (hunt-hunter) while others are not (e.g. corn-corner, cent-center, sock-soccer)

- Based on the English Lexicon Project (Batola et al. 2007)
  - **hunt-er** type: 94, **cent-er** type: 18
  - **-er** is productive: \(18 < 24 = 112 / \ln(112)\)
  - **hunt-er** is stored compositionally, and **center** is stored holistically (exception)
  - **-th** fails as a productive nominal suffix: warmth, width trumped by dearth, forth, tooth
Conception of Learning

\[ G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow \ldots \rightarrow G_T \]

- If a grammar fails to reach productivity as prescribed by TP, then it’s rejected.
- If there are multiple grammars meeting the TP threshold, then pick the grammar with fewest exceptions (i.e. more productive).
- If no grammar is productive, then the stress patterns of words will be memorized entirely listed.
- Learner processes the input data incrementally (in two stages).
- Ecological viability: the well-being of the child cannot depend on hearing *Manitoba* or *Winnipesaukee*. 
Conception of the Grammar

\[ G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow ... \rightarrow G_T \]

- There is a quasi-parametric system for stress and a finite set of inventories that could figure into stress computation and acquisition
  - e.g. quantity sensitivity, directionality in foot construction, position of projection
- These cannot all be innate but the learner must know the potentially relevant partition of the space (e.g., nouns and verbs)
- The sequence needn’t be innate (e.g., Drescher & Kaye 1990) but certain preferences can be encoded
- We evaluate two leading theories of stress (Halle & Vergnaud 1987, Halle 1998) for their explicitness, focusing only on primary stress placement
Conception of the learner

• That the learner has acquired the segmental inventory of her language (Werker & Tees 1983, Kuhl et al. 1992)

• That the learner has acquired the phonotactic constraints to facilitate syllable construction (Jusczyk et al. 1993)

• That the learner is capable of extracting words from continuous speech (Jusczyk & Aslin 1995, Yang 2004)

• That the learner is capable of detecting stress levels in words (Jusczyk et al. 1993)

• That the learner is capable of morphological segmentation at least in the inflectional domain (Brown 1973, Lignos & Chan 2011)
Conception of the data

- Child-directed English speech from CHILDES: 4.5M words and 27K types
- Used only nouns and verbs (automatically tagged, with morphological analysis): majority of words
- A combination of CELEX and CMUDICT to extract the stress contours of words: primary vs. non-primary
  - record vs. record
- Assume that universally, syllables with long vowels (diphthongs and /i/ and /u/) are heavy and short vowels without coda are light, and short vowels with coda can either be H or L and the child has to learn it.
Halle & Vergnaud (1987)

a. Nouns:
   - If the final syllable contains a long vowel (VV), it receives primary stress.
   - Otherwise if the penult is heavy (i.e., VV or VC\(^+\), short vowel with at least one consonant coda), then the penult receives primary stress.
   - Otherwise the antepenult receives primary stress.

b. Verbs:
   - If the final syllable is super heavy (i.e., VV or VCC\(^+\), a short vowel with at least two consonants in the coda), then the final syllable receives primary stress.
   - Otherwise the penult receives primary stress.

Halle (1998)

a. Nouns:
   - If the penult is heavy (i.e., VV or VC\(^+\)), then it receives primary stress.
   - Otherwise the antepenult receives primary stress.

b. Verbs: Same as HV87 above
An Initial Stage

Only words appearing at least once per 10K words

<table>
<thead>
<tr>
<th>contour</th>
<th>counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>287</td>
</tr>
<tr>
<td>10</td>
<td>107</td>
</tr>
<tr>
<td>100</td>
<td>13</td>
</tr>
<tr>
<td>01</td>
<td>7</td>
</tr>
<tr>
<td>010</td>
<td>3</td>
</tr>
<tr>
<td>1000</td>
<td>3</td>
</tr>
</tbody>
</table>

- 402 words: the QI system has 10 exceptions but can tolerate 402/ln(402) or 67 exceptions
- The initial stage of stress acquisition is QI in English (Kehoe & Stoel-Gammon 1997) and Dutch (Fikkert 1994); cf. Spanish (Hochberg 1988)
  - children often mis-stress (balloon - balloon)
  - With truncation (olifant - ofant)
- No reason to invoke quantity sensitivity as there are too few exceptions
A later stage

• Words at least once per Million

• 5763 words, requiring **5097** to be productive
  
  • trochaic: only **4533**
  
  • initial stress: only **4960** (86%)

• There is no evidence for QI system in the later stages of acquisition 
  (Stage 3, no later than 2;0: Fikkert 1994)
  
  • adult evidence for the role of syllable structures in stress placement (Guion et al. 2003, Baker & Smith 1976)

• Overwhelming majority does not guarantee productivity: filibuster proof required
Choices

- The learner needs to look at further partitions of the data
- Choice 1 (Lex): dividing words into categories (Nouns and Verbs)
- Choice 2 (Stem): remove inflectional endings, “beginning”→“begin”
  - do not assume the child can segment derivational morphology (Jarmulowicz 2002)
  - inflectional endings do not alter stress: trivial application of TP
Acquisition with Exceptions Succeeds

<table>
<thead>
<tr>
<th>lex</th>
<th>stem</th>
<th>HV$_{87}$</th>
<th>H$_{98}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>-</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>no</td>
<td>yes$^a$</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>no</td>
<td>yes$^b$</td>
</tr>
</tbody>
</table>

a. 515 exceptions, b. 355 exceptions
Kangaroo vs. Creamery

- HV87 treats kangaroo as the core case
- H98 treats creamery as the core case
- This empirical choice of H98 is borne out
- H98 and HV87 also use distinct theoretical apparatuses: notably, H98 allows bracket insertion rather than foot construction
  - These analytical choices are not crucial, so far as we can see
- Theoretical decisions have acquisition implications
Derivational Extensions

- Eventually children learn derivational morphology and their stress properties (Jarmulowicz 2002)
- Some suffix shift stress (-ation) while others do not (-ment) but many have exceptions (station-stationary, but document-documentary)
- -ary has been treated as non-shifting
- Using the English Lexicon Project and its morphological segmentation
- Using only words that could have shifted (i.e., not using tone-tonic)

<table>
<thead>
<tr>
<th>suffix</th>
<th>shifting</th>
<th>N</th>
<th>m</th>
<th>valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ment</td>
<td>no</td>
<td>201</td>
<td>0</td>
<td>yes</td>
</tr>
<tr>
<td>-ary</td>
<td>no</td>
<td>41</td>
<td>8</td>
<td>yes(^a)</td>
</tr>
</tbody>
</table>

\[ a. \frac{41}{\ln(41)} = 11 > 8 \]
Derivational Extensions

-ic and -ous have been treated as stress-shifting

What if the learner treats them as stress preserving?

<table>
<thead>
<tr>
<th>suffix</th>
<th>shifting</th>
<th>(N)</th>
<th>(m)</th>
<th>valid</th>
</tr>
</thead>
<tbody>
<tr>
<td>-ic</td>
<td>no</td>
<td>135</td>
<td>120</td>
<td>no</td>
</tr>
<tr>
<td>-ous</td>
<td>no</td>
<td>90</td>
<td>30</td>
<td>no(^a)</td>
</tr>
</tbody>
</table>

\[a. \frac{30}{90}/\ln(90)=20\]

Treating -ous as stress-shifting (i.e. not ignored), the noun/adjective system works perfectly.
An extension to syntax

- English is not a pro/topic-drop language but there is some level of subject omission (imperatives, diary drop, “looks fine to me”)
- Fitting a statistical distribution (Yang 2002) will not work as subject omission has structural restrictions
- Child-directed English: 1000 utterances, 702 containing a verb
  - 68 do not contain subjects (virtually all imperatives and echo questions)
  - \(68 < \frac{702}{\ln 702} = 107\)
- 249 distinct verb types, 39 may omit subject, still below threshold (45)
- The exception/productivity detection model may generalize across linguistic domains
Diacritics can be recognized across domains, which enables the grammar

Possible to have no productive rule

Bilingualism/Bidialectalism: loanwords on steroid, no different from the noun vs. verb split in English stress acquisition

Exceptions are not a problem unless an alternative theory has fewer exceptions (SPE, Preface)

One needn’t resort to a radically lexicalized theory of grammar

On General Learning and Evaluation Procedure:

“is not given a priori . . . Rather, an proposal concerning such a measure is an empirical hypothesis about the nature of language” (Aspects, p37)