Factors affecting lexical development by English speaking children

Junko Maekawa and Holly L. Storkel, University of Kansas

Introduction

- Young children learn new words rapidly and effortlessly.
- Fast mapping
- Phonotactical variables affecting lexical development:
  - Phonetic probability – likelihood of occurrence of a phoneme sequence
    - High: phoneme sequences that are common in a language (e.g., “coat”)
    - Low: phoneme sequences that are rare in a language (e.g., “watch”)
  - Neighborhood density – the number of words (i.e., neighbors) that differ by one phoneme from a target word
  - Dense: A word has many neighbors (e.g., “sit” – 36 neighbors)
  - Sparse: A word has few neighbors (e.g., “these” – 9 neighbors)
- Word frequency: number of occurrence of a word in a language
- Word length: number of phonemes in the transcription
- High probability words from dense neighborhoods are acquired more rapidly than low probability from sparse neighborhoods (Storkel, 2001, 2004; Storkel & Maekawa, in press).
- Shorter words that occur frequently in a language are acquired at earlier ages than longer words that occur infrequently (Storkel, 2004).

Questions

- Do phonotactic probability, density, word frequency, and/or word length predict the age when words are first produced spontaneously by young children?
- Are there similarities and differences across children in the patterns of significant predictors?

Method

- Conversational samples from CHILDES
- April: 22-35 months; 20 data points; 340 different root nouns
- Peter: 21-37 months; 20 data points; 767 different root nouns
- Allison: 16-34 months; 6 data points; 174 different root nouns

Variables

- Predictor variables
  - Phonetic probability (Segment average, Biphone average)
  - Density
  - Frequency (Kucera and Francis, 1967)
  - Length
- Outcome variable
- Age of first production
  - Earliest age when the child spontaneously produced the word in the sample

Analyses

- Primary analysis: backward regression to determine significant predictors
- Follow-up analyses: regression to determine predictors for each type of word for children with multiple significant predictors (i.e., Peter, Allison)

Results

- April:
  - Length: t = 4.22, p < 0.001, r² = 0.07
  - Short words were produced at earlier ages than long words
- Peter
  - Density: t = 2.11, p = 0.04, r² = 0.01
  - Words from sparse neighborhoods were produced at earlier ages than words from dense neighborhoods, contrary to the previous studies.
  - Frequency: t = -3.06, p = 0.002, r² = 0.03
  - High frequency words were produced at earlier ages than low frequency words
  - Length: t = 2.43, p = 0.02, r² = 0.02
  - Short words were produced at earlier ages than long words

Table 1. Results of the follow-up regression analyses by word types for Peter

<table>
<thead>
<tr>
<th></th>
<th>β estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse words</td>
<td>Frequency</td>
<td>0.10</td>
<td>0.03</td>
<td>-2.33</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.20</td>
<td>0.06</td>
<td>-1.96</td>
</tr>
<tr>
<td>Dense words</td>
<td>Frequency</td>
<td>0.90</td>
<td>0.06</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.23</td>
<td>0.07</td>
<td>-1.96</td>
</tr>
<tr>
<td>Low frequency words</td>
<td>Density</td>
<td>0.11</td>
<td>0.03</td>
<td>1.64</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.06</td>
<td>0.01</td>
<td>0.98</td>
</tr>
<tr>
<td>High frequency words</td>
<td>Density</td>
<td>0.11</td>
<td>0.06</td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td>0.20</td>
<td>0.05</td>
<td>3.00</td>
</tr>
<tr>
<td>Short words</td>
<td>Density</td>
<td>0.14</td>
<td>0.04</td>
<td>2.14</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>0.21</td>
<td>0.04</td>
<td>-1.87</td>
</tr>
<tr>
<td>Long words</td>
<td>Density</td>
<td>0.07</td>
<td>0.02</td>
<td>0.87</td>
</tr>
<tr>
<td></td>
<td>Frequency</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.09</td>
</tr>
</tbody>
</table>

- Allison
  - Segment average (Phonotactic probability): t = -2.02, p = 0.05, r² = 0.03
  - High probability words were produced at earlier ages than low probability words
  - Length: t = 2.89, p = 0.005, r² = 0.05
  - Short words were produced at earlier ages than long words

Table 2. Results of the follow-up regression analyses by word types for Allison

<table>
<thead>
<tr>
<th></th>
<th>β estimate</th>
<th>Standard Error</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low probability words</td>
<td>Length</td>
<td>0.15</td>
<td>0.08</td>
<td>1.84</td>
</tr>
<tr>
<td>High probability words</td>
<td>Length</td>
<td>0.28</td>
<td>0.09</td>
<td>3.11</td>
</tr>
<tr>
<td>Short words</td>
<td>Segment average</td>
<td>-0.12</td>
<td>0.07</td>
<td>-1.69</td>
</tr>
<tr>
<td>Long words</td>
<td>Segment average</td>
<td>-0.04</td>
<td>0.08</td>
<td>-0.55</td>
</tr>
</tbody>
</table>

Summary

Table 3. Summary of predictors for each child (with early acquired word types noted in parentheses) and of differences across children

<table>
<thead>
<tr>
<th></th>
<th>April</th>
<th>Peter</th>
<th>Allison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall analysis</td>
<td>Length (short)</td>
<td>Density (sparse)</td>
<td>Frequency (high)</td>
</tr>
<tr>
<td>Low probability</td>
<td>NA</td>
<td>NA</td>
<td>No significant predictors</td>
</tr>
<tr>
<td>High probability</td>
<td>NA</td>
<td>NA</td>
<td>No significant predictors</td>
</tr>
<tr>
<td>Sparse</td>
<td>Frequency (high)</td>
<td>Length (short)</td>
<td>NA</td>
</tr>
<tr>
<td>Dense</td>
<td>Frequency (high)</td>
<td>Length (short)</td>
<td>NA</td>
</tr>
<tr>
<td>Low frequency</td>
<td>NA</td>
<td>No significant predictors</td>
<td>NA</td>
</tr>
<tr>
<td>High frequency</td>
<td>NA</td>
<td>No significant predictors</td>
<td>NA</td>
</tr>
<tr>
<td>Short</td>
<td>Density (sparse)</td>
<td>Frequency (high)</td>
<td>No significant predictors</td>
</tr>
<tr>
<td>Long</td>
<td>No significant predictors</td>
<td>No significant predictors</td>
<td>No significant predictors</td>
</tr>
</tbody>
</table>

Discussion

- The predictors of age at first production varied across children and across word types
- Possible accounts for the observed variability
  - Lexical variability
    - Differences in size of the lexicon and frequency of words for each child did not predict the variability.
  - Use and availability of cues
  - Emergentist Coalition Model (Hirsh-Pasek, Golinkoff, & Hollich, 2000)
    - A range of cues are available for a child to learn words.
    - Availability of these cues varies over time.
    - Word length may be one of the earliest cues used to learn words.

Conclusion and future investigation

- Theories of lexical development need to be able to account for individual differences in the predictors of word learning across children, including other populations such as speakers of English as a second language, bilingual speakers, or monolingual Japanese children.

Reference


