Longitudinal Naturalistic Word-Learning by Toddlers

Junko Young
Holly Storkel
University of Kansas, Lawrence
Three Levels of Representations

Semantic (Concept / Referent)

Lexical (Whole word form)

Phonological (Sounds)
Lexical Representations

Organized into similarity neighborhoods
(Luce & Pisoni, 1998)

_similarity neighborhoods_: all the words differing by one phoneme substitution, deletion, or addition

kit:
sit, coat, king, it, skit
Dense vs. Sparse Neighborhoods

*Dense neighborhoods:* 
Words with many neighbors e.g., ‘sit’

*Sparse neighborhoods:* 
Words with few neighbors e.g., ‘these’
Word Learning

Previous Studies

- Empirical studies (Storkel, 2001; Storkel & Rogers, 2000)
  - 3 to 13-year-old children
  - Acquisition of novel words
  - Other factors are controlled
    Dense > Sparse

- Naturalistic word learning (Storkel, 2003)
  - Neighborhood Density ~ Age of Acquisition
    Dense > Sparse
Word Learning
Previous Studies -- Limitations

- Limits external validity of neighborhood density effects
- Focused on short term learning and retention
- Cross-sectional
Word Learning

Other factors

- Word Frequency
- Word Length
- Phonotactic Probability
Word Learning

Other factors

Word Frequency

- # of occurrences of a word in a language

- Positively correlated with neighborhood density (Landauer & Steeter, 1973)

- Kucera & Francis (1967)
Word Learning

Other factors

Word Length
- # of phonemes in the phonetic transcription

-Negatively correlated with neighborhood density (Bard & Shillcock, 1993; Pisoni et al., 1985)
Word Learning

Other factors

Phonotactic Probability

- The likelihood of occurrence of sound sequence (common vs. rare)
  - Positional Segment Average
  - Biphone Average

- Positively correlated with neighborhood density (Vitevitch et al., 1999)
Questions

- What factors affect long term learning and retention?
  - Neighborhood density?
  - Word Frequency?
  - Word Length?
  - Phonotactic Probability?

- Are these factors consistent across individual children?
Method

Language samples

- 3 children’s spontaneous speech samples (Higginson, 1985; Bloom, 1970; 1973) from the CHILDES database (MacWhinney, 2000)

- April: 22-35 months; 340 different nouns; 6 data points
- Peter: 21-37 months; 767 different nouns; 20 data points
- Allison: 16-34 months; 174 different nouns; 6 data points
Analysis

-Nouns (root)

-Backward Regression
  -Outcome Variable:
    -Age of Acquisition:
      -The earliest age that a word was spontaneously produced

-Predictor Variables:
  -Neighborhood Density
  -Word Frequency
  -Word Length
  -Phonotactic Probability
## Results

<table>
<thead>
<tr>
<th></th>
<th>Density</th>
<th>Frequency</th>
<th>Length</th>
<th>Seg avg</th>
<th>Biph avg</th>
</tr>
</thead>
<tbody>
<tr>
<td>April</td>
<td>ns</td>
<td>ns</td>
<td>$p&lt;.001$ ( + )</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Peter</td>
<td>$p=.035$ (+)</td>
<td>$p=.002$ ( - )</td>
<td>$p=.015$ ( + )</td>
<td>ns</td>
<td>ns</td>
</tr>
<tr>
<td>Allison</td>
<td>ns</td>
<td>$p=.08$ ( + )</td>
<td>$p=.005$ ( + )</td>
<td>$p=.045$ ( - )</td>
<td>ns</td>
</tr>
</tbody>
</table>
April

-Length
-Short words acquired earlier than long words
Peter

- Neighborhood Density
  - Sparse words acquired earlier than dense words

- Word Frequency
  - Frequent words acquired earlier than infrequent words

- Length
  - Short words acquired earlier than long words
Allison

- Word frequency
  - Infrequent words acquired earlier than frequent words

- Length
  - Short words acquired earlier than long words

- Positional segment average
  - Common words acquired earlier than rare words
Conclusion

1. Variety of factors affect long term learning and retention.
   1. Neighborhood density
   2. Word frequency
   3. Word length
   4. Positional segment average

2. Factors are not consistent across individuals.
Acknowledgements

NIH DC04781
ASHA Student Travel Award

Maki Sueto
Jennie Fox
Michael Vitevitch

Contact:
junko@ku.edu
http://www.ku.edu/~wrdlrng/