Lexical & Semantic Similarity in Word Learning

Holly L. Storkel, Ph. D.
Speech-Language-Hearing: Sciences & Disorders

Model of the Lexicon: Assume 2 Representations

- Lexical Representation
- Semantic Representation

Quantifying Lexical Similarity: Neighborhood Density
- Number of phonologically similar words
  - Differing by 1 sound
  - Sparse /dɛɡ/ (6) vs. dense /kæt/ (27)
- Influences recognition, production, & memory

Quantifying Semantic Similarity: Multiple Methods
- Number of meaningfully related words
  - Small (cat - 3) vs. large (bird - 23)
- Methods vary in how similarity determined
  - Co-occurrence
  - Feature analysis
  - Association data
- Influences memory
  - Emerging studies of recognition
Similarity Matters for Known Word Retrieval

Step 2: Configuration

Step 3: Engagement

What about word learning?

/ɡουm/

Step 1: Triggering

Question

Does similarity to existing representations influence word learning?
Overall Purpose

• How to determine similarity for novel words?
  – Study 1: Neighborhood density
  – Study 2: Set size

• Does similarity influence word learning?
  – Study 3: Neighborhood density
  – Study 4: Set size

Study 1: Questions

• How do child and adult neighborhood density values compare?
  – Methodological issues: Do we need child values?
  – Theoretical issues: How does the lexicon grow from child to adult?

Study 1: Neighborhood Density (Storkel, Hoover, & Kiewig)

• Relatively well established density algorithm
  – Count the number of words in a corpus that differ by one phoneme
  – Works for real words and novel words

• Question: What corpus?

Growth of the Lexicon

• Hypothesis 1:
  – Words added to lexicon are equally distributed across neighborhoods
    • Number of neighbors increases equally for all words

<table>
<thead>
<tr>
<th></th>
<th>Child Density</th>
<th>Adult Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>bird</td>
<td>20</td>
<td>23</td>
</tr>
</tbody>
</table>

Adult Corpus

• Hoosier Mental Lexicon (HML)
• Dictionary of ~20,000 words
• Computer readable phonemic transcription
• Used in most past research of children & adults
  – Is this appropriate?

Growth of the Lexicon

• Hypothesis 2:
  – Words added to lexicon are asymmetrically distributed across neighborhoods
    • Number of neighbors increases differently for different words

<table>
<thead>
<tr>
<th></th>
<th>Child Density</th>
<th>Adult Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>bird</td>
<td>13</td>
<td>23</td>
</tr>
</tbody>
</table>
Methods

- Set of real words
- Compute two neighborhood densities
  - One based on adult corpus (HML)
  - One based on a child corpus (next slide)
- For both computations (adult & child), use 1 phoneme change algorithm

Child Corpus

- Based on spoken data from Kindergarten and 1st grade children
  - Kolson (1960)
- Calculator created by Storkel, Hoover, & Kiewig
  - Provided phonemic transcriptions using HML conventions
  - Eliminated homonyms and morphologically related words
    - ~5,000 words
  - Calculator interface: [http://www2.ku.edu/~wrdlrg/]
Results: Correlation

Results: T-test
• Adult density values were significantly larger than child density values
• This is the expected pattern if the lexicon is growing

Asymmetric Growth

Study 1 Conclusion
• Adult and child density values highly correlated
  – Words that are dense in one corpus will tend to be dense in the other
• However, growth is asymmetric
  – Few words added to sparse neighborhoods
  – Many words added to dense neighborhoods
  – Difference between sparse and dense increases

Question
Is the growth equivalent or asymmetric?
Study 2: Set Size  
(Storkel & Adlof, in press a)

- How to determine semantic similarity for nonobjects?
- How do child and adult values compare?  
  - Methodological issues: Do we need child values?  
  - Theoretical issues: How does the lexicon grow from child to adult?

Determining Semantic Similarity

- Many methods used with real objects won’t work for nonobjects  
  - Ex. – frequency of co-occurrence
- Only viable method = discrete association  
  - Developed by Nelson, McEvoy, & Schreiber for real words

Our Study

- Participants  
  - 82 undergraduate college students  
  - 92 preschool children
- Stimuli  
  - Nonobjects from Kroll & Potter (1984)  
  - Only 47 of 88 tested on adults & children

Question

How do child and adult set sizes compare?

Discrete Association Method

- Show picture (or word)  
- Participant reports first word that comes to mind
- Word reported by two or more participants = neighbor  
- Number of neighbors = set size

Results: Correlation

![Graph showing correlation between child and adult set sizes.](image-url)
Results: T-test

- No significant difference
- Adult values tended to be similar to child values
  - No growth?!?
- Actual words reported as neighbors differed across children and adults (see article)

---

Study 2 Conclusion

- Adult and child set size values similar
  - (Although actual semantic neighbors differ)
- No semantic growth?
  - Unlikely
  - Probably an artifact of the stimuli (i.e., nonobjects)
  - Need to examine real words to determine growth pattern

---

Study 3 & 4: Basic Word Learning Method

- Participants
  - Typically developing preschool children
- Stimuli
  - Nonwords paired with nonobjects
  - Manipulate one variable
  - Hold other variables constant
- Exposure in game format on computer
- Test referent identification & naming
  - Prior to exposure (baseline)
  - End of exposure (learning)
  - 1-week post-exposure (retention)

---

Exposure Example

This is a goUm. Say goUm. [goUm]. Yes, that's a goUm. Remember, it's a goUm. We're going to play a game. Find the goUm. That's the goUm. Say goUm. [goUm]. Don't forget the goUm.

---

Question

Do these variables influence word learning?
Study 3 & 4

---

Naming Example

goUm
Referent Identification Example

Stimuli Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Density</td>
<td>5 (1)</td>
<td>12 (2)</td>
</tr>
<tr>
<td>Semantic Set</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>10 (3)</td>
<td>11 (3)</td>
</tr>
</tbody>
</table>

Study 3: Neighborhood Density (Storkel & Rosales, in progress)

- Participants: 25 preschool children
- Vary neighborhood density of nonwords
  - Low vs. High
- Hold semantic set size of nonobjects constant
  - Mid

Question

Does neighborhood density influence word learning?

Stimuli

Yes: Sparse > Dense
Study 3 Summary

• Sparse learned better than dense
  – Only at criterion test (i.e., immediate learning)
  – Not at retention test

Stimuli Characteristics

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighborhood Density</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Semantic Set Size</td>
<td>8</td>
<td>13</td>
</tr>
</tbody>
</table>

Study 4: Set Size
(Storkel & Adlof, in press b)

• Participants: 36 preschool children
• Hold neighborhood density of nonwords constant
  – Mid
• Vary semantic set size of nonobjects
  – Low vs. High

Stimuli

<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>goum</td>
<td>boug</td>
<td></td>
</tr>
<tr>
<td>mafi</td>
<td>wun</td>
<td></td>
</tr>
<tr>
<td>hsg</td>
<td>pig</td>
<td></td>
</tr>
<tr>
<td>jem</td>
<td>jat</td>
<td></td>
</tr>
<tr>
<td>fip</td>
<td>metb</td>
<td></td>
</tr>
</tbody>
</table>

Question

Does semantic set size influence word learning?

Yes: Small > Large

Referent Identification:
- Small Set Size
- Large Set Size

Naming:
- Small Set Size
- Large Set Size

Referent Identification: Small Set Size
Baseline
Criterion
Retention

Referent Identification: Large Set Size
Baseline
Criterion
Retention

Naming: Small Set Size
Baseline
Criterion
Retention

Naming: Large Set Size
Baseline
Criterion
Retention

* *
Study 4 Summary

• Small set size learned better than large set size
  – Not at criterion test (i.e., immediate learning)
  – Only at retention test

Study 3 & 4 Summary

• Lexical and semantic similarity both influence learning
  – Less similar items learned better than more similar

• Timing of influence varies
  – Lexical = immediate learning
  – Semantic = retention

Hypothesis 1

• Lexical similarity influences early components of word learning
  – Triggering of learning
    • Low similarity = more unique
    • More obvious that lexical representation does not exist for new word
    • Learning immediately triggered

Step 1: Triggering

Step 3: Engagement
Hypothesis 2

- Lexical and semantic similarity both affect engagement but timing of engagement varies
  - Early engagement for lexical representations
    - Overlap in word learning “steps”
  - Late engagement for semantic representations
    - Less overlap in word learning steps

Step 3: Engagement

Future Directions

- Identify measures that are sensitive to
  - Triggering
  - Configuration
  - Engagement

- Test the role of lexical and semantic similarity in each process

Acknowledgements

DC 008095 (Storkel)
DC 009133 (Adlof)
DC 009135 (Hoover)
DC 000052 (T-Grant)
DC 005803 (BNCD)

Staff of the Word & Sound Learning Lab

DC 008095 (Storkel)
DC 009133 (Adlof)
DC 009135 (Hoover)
DC 000052 (T-Grant)
DC 005803 (BNCD)

Staff of the Word & Sound Learning Lab