Predictive modeling of language acquisition using network growth models

Nicole Beckage and Eliana Colunga
University of Kansas
University of Colorado, Boulder
Child language: a case study

Environment and language input

Interests of child

Learning

Knowledge

Observed: Productive vocabulary
Colunga Lab collected longitudinal data, following specific children for about a year.

- 114 children with longitudinal CDIs (L-CDIs).
- On average 9.4 CDIs per child.
- About 1000 vocabulary learning snapshots.
Focus on words learned next

- What are the cognitive mechanisms and developmental processes that support future language learning?
  - Can we construct computational models that predict future acquisition of *individual* words?
Network growth models allow for exploration of cognitive theory.

Algorithmic assumptions provide formalism to attentional mechanisms of learning and acquisition.
Nodes: words that the child produces.

Edges: Predefined relationship between words. Here, free associations.

Edges are directed: Bird elicits response of tree, but tree does not elicit bird.
Networks: levels of representation

- **Macro level**: define a relationship between nodes
  - Defining a graph.
- **Mezzo level**: measure change in structure
  - Defining a process.
- **Micro level**: study interaction between nodes
  - Defining a measure of node importance.
Network framework

- Define a graph’s edge list.
- Define a process.
- Define a measure of word importance.

Hypothesis:
If we have the right definition of *relationship*, the right growth *process*, and the right *measure* of importance then we can accurately predict words to be learned next.
Macro level:

- Define a graph’s edge list.
- Define a process.
- Define a measure of word importance.
- Hypothesis:
  We can measure what relationships are meaningful to a child.
How do the sounds of words (phonology) and meaning of words (semantics) influence learning?

- Experimental manipulations with toddlers is difficult.
- Chicken and the egg problem…
  If one is more important, is it driving acquisition?
  Or does it appear important based on acquisition trends?

Network construction:

- Semantics: sensory-motor features, association norms.
- Phonological: edit distance one, phonemic overlap.
Mezzo level:

- Define a graph’s edges.
- Define a process.
- Define a measure of word importance.
- Hypothesis:
  We can measure the process of changing structure in relation to the learner and environment.
Growth processes:

Mezzo level includes interaction

- of current vocabulary,
- word to be learned,
- connectivity in full vocabulary (environment).

Is the driving force from the environment or from the child’s current vocabulary?

- Altering the effect of each allows us to capture the role of the internal and external environment.
- Consider the interaction of growth given semantic vs. phonological influences.
Model A: Preferential attachment

- Attractor model on subset of known words.
- **Known words** with a high number of connections are attractors.
- Unknown words learned if they attach to highly connected words.
- Subset of **known words** plays largest role in acquisition predictions.

### Preferential-Attachment Model

<table>
<thead>
<tr>
<th>Node</th>
<th>Value</th>
<th>Add</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>3</td>
<td>+</td>
</tr>
<tr>
<td>N2</td>
<td>0</td>
<td>−</td>
</tr>
<tr>
<td>N3</td>
<td>0</td>
<td>−</td>
</tr>
</tbody>
</table>
Model A: Preferential attachment

- Attractor model on subset of known words.
- **Known words** with a high number of connections are attractors.
- Unknown words learned if they attach to **highly connected words**.
- **Subset of known words** plays largest role in acquisition predictions.
Model B: Preferential acquisition

- Attractor model on subset of unknown words.
- **Unknown words** with high number of connections are attractors.
- Unknown words learned if highly connected in **learning environment**.
- **Independent** of words the child knows; **language environment** drives learning.

<table>
<thead>
<tr>
<th>Node</th>
<th>Value</th>
<th>Add</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>N2</td>
<td>2</td>
<td>–</td>
</tr>
<tr>
<td>N3</td>
<td>3</td>
<td>+</td>
</tr>
</tbody>
</table>
Model B: Preferential acquisition

- Attractor model on subset of unknown words.
- **Unknown words** with high number of connections are attractors.
- Unknown words learned if highly connected in **learning environment**.
- **Independent** of words the child knows; **language environment** drives learning.
Model C: Lure of the associates

- Intersection of previous models.
- Attractor model on unknown words counting edges coming from subset of known words only.
- Combines the role of the language environment and the child’s current vocabulary.

<table>
<thead>
<tr>
<th>Node</th>
<th>Value</th>
<th>Add</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>N2</td>
<td>2</td>
<td>+</td>
</tr>
<tr>
<td>N3</td>
<td>1</td>
<td>–</td>
</tr>
</tbody>
</table>
Model C: **Lure of the associates**

- Intersection of previous models.
- Attractor model on **unknown** words counting edges coming from subset of **known words** only.
- Combines the role of the **language environment** and the child’s **current vocabulary**.
Levels of representations

- Define a graph’s edges.
- Define a process.
- Define a measure of word importance.
Micro level:

- Define a graph’s edges.
- Define a process.
- Define a measure of word importance.
  - Hypothesis: We can explore measures that help explain the learning of new words.
Defining centrality measures

\[ Pr(w_i) = \frac{e^{\beta d_i}}{\sum_j e^{\beta d_j}} \]

\[ d_i = \sum_{j \in V} I_{(i,j) \in E} \]

**Degree centrality**: Here \( d_i \) is a function of the number of edges coming from word \( w_i \) to any other word in the graph \( G(V, E) \).

\[ d_i = \sum_{h \neq i \neq j} \frac{\sigma_{hj}(i)}{\sigma_{hj}} \]

**Betweenness centrality**: Here \( d_i \) is a function of the number of shortest paths \( \sigma_{hj} \) between words \( h \) and \( j \) that contain word \( w_i \) in the graph \( G(V, E) \).
Defining centrality measures

\[ \Pr(w_i) = \frac{e^{\beta d_i}}{\sum_j e^{\beta d_j}} \]

\[ d_i = \sum_{j \in V} I_{(i,j) \in E} \]

**Degree centrality:** What’s cognitively important is how many relationships a word has.

\[ d_i = \sum_{h \neq i \neq j} \frac{\sigma_{hj}(i)}{\sigma_{hj}} \]

**Betweenness centrality:** What’s cognitively important is that new words shorten the distance between known words.
Results: Predictive models

- Different **centrality measures** result in different distributions of best fitting models.

  - The best fitting model for **semantics** is **preferential acquisition** using **degree** centrality.

  - The best fitting model for **phonology** is **lure of the associates** using **betweenness** centrality.
Results: Predictive models

- Different **centrality measures** result in different distributions of best fitting models.

- The dominant mechanism for **semantic learning** is connectivity in **learning environment**.

- The dominant mechanism for **phonological learning** is related to learning words that bridge phonological clusters.
Conclusions: Language acquisition

Network analysis allows us to formally model

- whether **semantics** or **phonology** is most salient.
- the role of the **current vocabulary** and the **language environment** on learning future words.
- operationalize linguistic **importance** of early words.

- Potential to build more complete model of the **acquisition process** and how that process changes over development.
Conclusions: Network framework

Network analysis allows us to formally model the

- Types of relationships that are relevant.
- Process of growth operating on the changing network.
- The role of individual nodes in altering structure.

- We can test implications and support for understanding learning and cognition through algorithmic assumptions.
  - Educational and learning situations.
  - Social and behavioral models.
Future work

- We know there are populations of learners, do learners vary systematically at any of these levels?
  - Categorize and explain differences simultaneously.
  - Understand ways to correct attention and input.

- These network growth models are new from an algorithmic perspective.
  - What other types of questions can these tools address?
    - Educational modeling, social dynamics.
  - Algorithms and techniques allow for changing nodes.
Thank you!

Email: beckage@ku.edu

Special thanks to Eliana Colunga, the DACS lab, Michael Mozer, and Thomas Hills.