Network Growth Models
Applied to Language Acquisition

Nicole Beckage and Eliana Colunga
University of Colorado, Boulder
Lexical Acquisition: an overview

- Environment and language input
- Interests of child

Learning

Knowledge

Observed: Productive vocabulary

Nicole Beckage
Motivation for modelling acquisition

- Modeling offers a way of **testing** existing hypotheses on language acquisition.

- We can also use modeling results to help **inform** experimental and theoretical work.

- If we better understand the **process** of acquisition
  - We can better **diagnose** language learning difficulties
  - And potentially provide tools to help **aid** in language learning

- **Goal:** Predict what words a child is likely to learn next given their current vocabulary.
Vocabulary data

MacArthur Bates Communicative Development Inventory (CDI, MCDI)

- Closed-form checklist of nearly 700 words.
- Parents report words their child can *produce*.
- Normative age of acquisition information.
Age 18mo
Levels of Representations

- **Macro level**: what defines a relationship between words
  - Defining a *graph*

- **Mezzo level**: interaction of nodes with representations
  - Defining a *process*

- **Micro level**: interaction of nodes with other 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 edges, the right 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.
Language is relational

- What are the types of relationships that matter for learning?
- Does the importance of representation change over the course of development?

Simple initial question:

- **Semantic** network (sensory motor features, Howell 2005)
- **Phonological** network (Phonemic overlap)

What aspects of language are salient and important from the perspective of a child
Macro model:

- Assume edge list.
- Start with current vocabulary. Given a process, probabilistically select words for learning.
- Compute percent overlap between predicted (modeled) and learned (reported monthly by parents).
- Empirical estimate (average over 1000 runs).
- Aggregate snapshots by age, percentile and vocabulary size.
Macro model:

Define: $G(V,E)$ to be the complete vocabulary graph

$G^*(V^*,E^*)$ be based on the child’s current vocabulary

Note $G^*(V^*,E^*) \subseteq G(V,E)$

Assume that children learn words according to a preferential mechanism ‘rich get richer’:

$$\Pr(\text{learn}(j)) \propto \sum_{i \in V^*, j \notin V^*} d_i I_{\{(i,j) \in E\}}$$

Where $d_i$ is the degree or edges that word $i$ has in graph $G^*$.

That is, a word is likely to be learned if it connects to other highly connected known words.
Results: Macro analysis

- Is a **phonological** or **semantic** network representation better assuming preferential growth?
We can measure how representational assumptions affect modeling results.

For acquisition we see:

- Early in learning, and for slower language learners, **random selection** is hard to beat.
- Later in learning and for faster language learners, **phonological information** seems to be more available.

This is for **normative** vocabularies... future work – individual trajectories.
Mezzo level:

- Define a graph’s edges.
- Define a process.
- Define a measure of word importance.
- Hypothesis: We can measure the role of the child’s knowledge and environmental information in learning.
Mezzo model:

- Fix semantic/phonological graph.
- Start with current vocabulary. Test which of three models predicts future acquisition.
- Compute empirical log-likelihood for withheld test set for both trajectories and snapshots.
- Consider effect of network and process on prediction accuracy.

Hills et al. 2009, 2010
Growth processes: Hills 2009, 2010

Mezzo level includes interaction

- of current vocabulary
- word to be learned
- future vocabulary growth

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

- Altering the roles of each allows us to capture the role of the child and the role of the language environment

- And the interaction of these roles with the definition of relationship
Colunga Lab collected longitudinal data, following specific children for about a year.

- 114 children with longitudinal CDIs (L-CDIs)
- On average 9 CDIs per child
- 905 vocabulary learning snapshots
Growth processes from past work

Models and image from Hills et al. 2009.
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

<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

Hills et al. 2009, 2010
Model B: Preferential acquisition

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

<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 a high number of connections are attractors
- Unknown words learned if they are highly connected in language learning environment
- Independent of words child knows
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 learning process

<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>

Hills et al. 2009, 2010
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 learning process

Hills et al. 2009, 2010
Results: Mezzo analysis

Compare **semantic** (free association norms, Nelson 2005) and **phonological** (Levenshtein distance, Vitevitch 2008) representations.

- **Preferential attachment** is the best in the **phonological** domain
- **Preferential acquisition** is the best in the **semantic** domain
Results: Mezzo analysis

Compare **semantic** (free association norms, Nelson 2005) and **phonological** (Levenshtein distance, Vitevitch 2008) representations.

- **Preferential attachment** is the best in the **phonological** domain
- **Preferential acquisition** is the best in the **semantic** domain

**Proportion of children best fit by each model**

- **Pref Att**
- **Pref Acq**
- **Lure**
Summary: Mezzo analysis

Which growth model performs best is related to what network representation is used.

There are process differences based on which representation is chosen.

The role of the current and future vocabulary can be considered in relation to learning.

These can be interpreted in terms of cognitive theory and cognitive assumptions.
Micro level:

- Define a graph’s edges.
- Define a process.
- Define a measure of word importance.
- Hypothesis: We can define and explore reasons for the learning of new words.
Micro model:

- How we define both the edges $E$ and the centrality measure used to define $d_i$ forms the acquisition model.

Hypothesis:
- if we have the right definition of edges $E$ and the right measure of centrality to compute $d_i$ we can predict words to be learned next.

- We explore edges, growth models and centrality in terms of fit and cognitive interpretation
Micro level: defining the relevant network statistics

Take a vocabulary network $G^*(V^*, E^*)$

Define the probability of learning a specific word $w_i$ to be

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

Where $d_i$ is a network centrality measure computed based on a network representation and $\beta$ is a sensitivity parameter.

Hills et al. 2009, 2010
Defining centrality measures

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

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

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

**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 \( G(V, E) \).

**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

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

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

**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.
Growth processes: degree

Models and image from Hills et al. 2009.
Different **centrality measures** result in different distributions of best fitting models.
Results: Defining Centrality

- Performance in terms of log-likelihood has no clear centrality measure.

![Graphs showing Degree, Eigenvector, and Betweenness Centrality](image.png)
Results: Defining Centrality

- The success of the growth process to generalize to new data depends on what measure of centrality is chosen.

- The model fit on unseen data does not widely vary between the types of centrality.

- 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.
Summary: Micro analysis

There is an interaction of centrality measure and network growth models.

There are assumptions about relevant information imbedded in the definition of ‘node importance’.

Considering different types of centrality allows for different attentional and cognitive implications to be explored.
Methodological summary

Define a graph’s edges.
- We can measure what relationships are meaningful to a child.

Define a process.
- We can measure the role of the child’s knowledge and environmental information in learning.

Define a measure of word importance.
- We can define and explore reasons for the learning of new words.
Levels of Representations

- Macro level:
  - Quantify relationships that matter to young children during early learning

- Mezzo level:
  - Consider the role of the learner and environment in learning

- Micro level:
  - Formalize what learning is trying to optimize
Conclusions

Network analysis allows us to formally model the

- Types of relationships relevant to learning
- Process of growth operating on the changing network
- The role of individual words in altering language structure

- Potential to combine all of these components both within and across levels of analysis.

- We can understand implications and support for cognitive theories
Extensions

- 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

- Vocabulary growth depends on both semantics and phonology, model jointly through
  - Edge lists capturing both relationships
  - Processes that weight each type of relationship
  - Node importance that is a function of both
Thank you!

Special thanks to Eliana Colunga
University of Colorado Boulder’s DACS lab
and Michael Mozer.

Questions?