Research Article

Network Growth Modeling to Capture Individual Lexical Learning

Nicole M. Beckage1 and Eliana Colunga2

1University of Wisconsin, Madison, USA
2University of Colorado, Boulder, USA

Correspondence should be addressed to Nicole M. Beckage; nicolebeckage@gmail.com

Received 8 September 2018; Revised 22 February 2019; Accepted 25 July 2019; Published 31 October 2019

Copyright © 2019 Nicole M. Beckage and Eliana Colunga. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Network models of language provide a systematic way of linking cognitive processes to the structure and connectivity of language. Using network growth models to capture learning, we focus on the study of the emergence of complexity in early language learners. Specifically, we capture the emergent structure of young toddler’s vocabularies through network growth models assuming underlying knowledge representations of semantic and phonological networks. In construction and analyses of these network growth models, we explore whether phonological or semantic relationships between words play a larger role in predicting network growth as these young learners add new words to their lexicon. We also examine how the importance of these semantic and phonological representations changes during the course of development. We propose a novel and significant theoretical framework for network growth models of acquisition and test the ability of these models to predict what words a specific child is likely to learn approximately one month in the future. We find that which acquisition model best fits is influenced by the underlying network representation, the assumed process of growth, and the network centrality measure used to relate the cognitive underpinnings of acquisition to network growth. The joint importance of representation, process, and the contribution of individual words to the predictive accuracy of the network model highlights the complex and multifaceted nature of early acquisition, provides new tools, and suggests experimental hypotheses for studying lexical acquisition.

1. Introduction

Children do not learn words in isolation. Instead children must learn the meanings and relationships of words within a communicative context, and in the context of other words. These same relationships, which make learning initial words challenging, likely offer scaffolding and support that help children make sense of the world around them. The connections and relationships between words likely aid future learning of new words. How the structure of language develops through the course of acquisition, and how this structure facilitates future language learning, is critical to understanding the acquisition process of early learners.

Here we set forth to build a predictive network growth model of the words a child is likely to learn next based on the emerging structure of the child’s current vocabulary. We represent the child’s current productive vocabulary as a network, with the words the child produces as nodes in the graph, and edges connecting words based on either semantic or phonological similarity. Our growth models assume that words enter the graph (become part of the child’s productive vocabulary) based on either the child’s current vocabulary knowledge or the structure of the global language environment as captured by the full network structure. With these assumptions, we have a systematic way of linking possible learning mechanisms and processes to the structure and connectivity of a child’s lexical network.

Although it is unlikely that language is represented in the mind as a network, it is probable that the structure of a child’s current vocabulary, or the structure of the language learning environment, influences language learning and lexical acquisition. Network representations provide a method for abstracting the complexity of language, allowing researchers to study the emergence of linguistic structure [1–4]. Here we make the simplifying assumption that language can be represented as a network and that we can learn
about the underlying cognitive process of acquisition by studying the change and growth of these language representations.

There is a rich amount of work considering both semantic and phonological networks of the adult lexicon (for a review, see [4–6]). Here we instead focus our review of previous work only on network analysis as related to the growing lexicon of early language learners. Structurally, early language networks show evidence of small-world structure, characterized by short average path lengths and high local clustering, even for very small graphs [1, 3, 7]. Interestingly, the small-world structure of these early lexical graphs has been shown to be correlated with language learning skills. In particular, 2-year-olds classified as late talkers, or those who have relatively small vocabularies compared to their peers, have been found to have lexical networks with less small-world structure than we would expect based on random network models [3]. This and other results suggest that the small-world structure of early language graphs is able to highlight or accentuate relevant features that may be important and facilitatory to future language learning, and thus small-world structure as observed in the growing lexical graphs may relate to the child’s language learning ability.

Previous work has also linked topological features of language networks to the process of acquisition. Steyvers and Tenenbaum [1] proposed that language is learned by a process similar to preferential attachment, which we call Preferential Growth to remind the reader that the assumed mechanisms and process of growth vary from the formalization of preferential attachment [8], with highly connected nodes being learned earliest and those highly connected nodes facilitating the learning of new nodes based on their connectivity to these early learned highly connected nodes. The model, which simulates aspects of semantic differentiation, suggests that words are more likely to be learned if they connect to already known, well-connected words in the child’s current lexical network [1]. Under semantic differentiation, new words or concepts are learned in relation to already known words [9–11]. In a network growth framework, a learned word attaches to highly connected nodes in the current vocabulary graph and then forms edges with the neighbors of the attachment node. The resulting network maintains the scale-free structure found in some semantic networks and also maintains high local clustering found in all language networks. This model suggests a mechanism for the observed empirical result of a strong correlation between the connectivity of words in semantic graphs and the reported age of acquisition of words [1].

Hills and colleagues [2, 7] suggest, instead, that language learning is driven by contextual diversity, or, in network terms, by the connectivity of unknown words in the language environment or the language of adults. Encoding adult language as a graph, connectivity of words may be related to structure in the environment (e.g., ball and catch), close proximity in spoken language (e.g., cat and dog), or even close proximity in physical space (e.g., chair and table). The connectivity of individual words in the full language graph may approximate the number of contexts and meanings of an individual word. Under a contextual diversity hypothesis, a word is more likely to be learned if it appears in multiple contexts since, with multiple exposures, the ambiguity of meaning will decrease [12, 13]. Experimental work has shown many ways in which multiple contexts and exposures can increase the likelihood of learning, as shown in cross-situational learning tasks [13], or via attentional mechanisms [14]. The model which operationalizes contextual diversity in network growth models is known as Preferential Acquisition and was proposed and validated in work by Hills and colleagues [2, 7, 15].

Here we propose a network modeling framework to predict the individual words a child is likely to learn next. While previous models have focused on normative acquisition [1, 2], we apply these growth models to the language trajectories of individual children. We additionally formalize a mathematical relationship between network analysis models of growth and the process of language acquisition in young children. The use of network models to explain acquisition requires (1) a clear definition of edges or similarity between words, (2) a systematic influence of the network structure on future vocabulary acquisition, and (3) a way of relating network structure to the acquisition of individual words. In the next section we formalize our framework for modeling individual acquisition before evaluating performance of the proposed models in the context of our framework. We find that each assumption (the graph, growth process, and conversion of network models to measures of probability of learning a specific word) plays an important role, impacting our ability to accurately capture the language acquisition of individual children. We argue that this framework offers novel insight into the acquisition processes involved in early word learning. Most promising is that our models outperform baseline especially when predicting the vocabulary development trajectories of children who are learning language slower than their peers. We find a strong relationship between this improvement in performance and assumptions about growth mechanisms underlying our model. In future work, we aim to explore whether the predictive accuracy of these network models can motivate intervention and/or empirical investigations, providing novel insight into the processes that underlie language acquisition in young children.

2. Network Growth Modeling Framework

When explaining vocabulary acquisition with network growth models, we should consider all the aspects of the network structure that might influence predictability and interpretability of the acquisition process. Toward this goal, we propose the following three levels of analysis to frame and understand our models and their ability to capture individual learning trajectories:

(1) Macro level: the definition of a graph in terms of nodes and edges. Edges are based on measures of relatedness that may be thresholded by a criterion specifying when an edge exists between two nodes,
(2) Mezzo level: measure of (changing) structure and influence of structural properties of a chosen graph on the topic of study.

(3) Micro level: the interaction of nodes with other nodes. Here we investigated different centrality measures as a proxy for node importance.

In the case of child language acquisition, the strongest form of the hypothesis assumes that if we have (1) a meaningful definition of relationships between words, (2) a growth process that correctly approximates learning, and (3) a cognitively relevant measure of word importance, then we can accurately model the acquisition process. We evaluate our assumptions on the model's ability to predict the specific words an individual child is likely to learn next given the child's current language network.

Beyond the goal of accurate prediction, we can also investigate performance of our models at each of these levels of analysis. For example, at the macro level, we can ask, given the acquisition trajectories of individual children, if semantic or phonological features are more predictive in capturing the acquisition trajectory of new words. The difference between phonological similarity and semantic feature similarity has previously been considered, assuming that words are added based on a "rich get richer" or preferential growth model [16], but under this framework, we can extend these results by exploring the interaction of phonology and semantic features with various growth processes, capturing performance of the mezzo level. We also examine cognitive processes at the micro level by asking such questions as whether words that bridge the network (have high betweenness) or words that have many connections (high degree) are more likely to be learned earlier.

By considering each of these levels of analysis, we can better understand the interactions among these three levels and their effect on explaining the acquisition trends of young children. We evaluate the role of each of these levels in predictive modeling and use these results to inform our understanding of the processes that influence early acquisition. These network models are not only predictive models but may also suggest mechanisms and attentional influences that alter individual language trajectories. To preview our results, we find that all levels of analysis matter, but that there is a high amount of variability in ways in which children learn, suggesting that future research must focus on understanding the relationship between the network structure and individual cognitive and developmental differences of young children.

3. Methods

We define the probability of a word \( i \) being learned by child \( x \) as

\[
Pr(i | x) = \frac{1}{1 + \exp \left( -\beta_0 + \beta_1 \delta_{\text{cdi}(i,x)} + \beta_2 \delta_{\text{la}(i,x)} \right)}
\]  

(1)

which is a logistic transformation that includes an intercept term (\( \beta_0 \)) and includes what we call the baseline word learnability for word \( i \) given child \( x \) (\( \delta_{\text{cdi}(i,x)} \)) and a model based measure or growth value for word \( i \) (\( \delta_{\text{la}(i,x)} \)), conditioned on the vocabulary of child \( x \). The model-based growth value is calculated based on network growth assumptions explained below. \( \beta \) s are free parameters that are learned in the training and validation portion of our models through standard logistic regression methods. In all models, we include \( \delta_{\text{cdi}(i,x)} \) which is the base rate probability that word \( i \) is learned according to the CDI norms, renormalized for the known vocabulary of the child \( x \) and constrained by the underlying network representation used to compute the growth value. If child \( x \) already produces a word, we do not compute the probability of learning that word.

Under our network growth framework used to predict the probability of word \( i \) being learned by child \( x \), the growth value \( \delta_{\text{la}(i,x)} \) is defined based on an assumed growth process and an assumed centrality (defined for three models in equations (3) through (5) which is naturally dependent on the underlying network representation).

The centrality measure can be either local (e.g., degree as this considers only immediate neighbors) or global (e.g., betweenness as this weighs the contribution of all nodes in the graph). The model-dependent growth value \( \delta_{\text{la}(i,x)} \), which is minimally zero, is mapped to a probability through a logistic transformation characterized by scaling parameters \( \beta \). Note that the above model allows for simultaneously predicting the set of words learned by an individual child, not only the single most likely word. Modeling learning of the single most likely word has been examined in a similar paradigm by Hills and colleagues using a mathematical formalism very similar to the model we propose here [2, 7]. \( \beta \) s are optimized across training data for each network representation, growth process, and centrality we compared and evaluated on unseen children and their vocabulary growth trajectory. \( \delta_{\text{cdi}(i,x)} \) is computed the same way for each model and varies for individual children only based on renormalization of the CDI norms for that specific child \( x \), as well as based on the words in the network representation used to compute \( \delta_{\text{la}(i,x)} \). We discuss \( \delta_{\text{cdi}(i,x)} \) and define the growth values for each model in more detail below.

In our modeling framework, we focus on a subset of early learned words that is widely studied in the developmental literature. These words are part of a language development checklist, the MacArthur-Bates Communicative Development Inventory, which contains words that at least 50% of children know by the time they turn 2 years (e.g., CDIs [17]). This checklist is also normed such that we know, on average, what percent of children at a specific age produce each word. Although we further discuss the words on this list—and research using these words—below, we note that using a fixed set of words poses a unique inference challenge for our modeling approach. We need to simplify the full language graph to the words on the CDI. But some of the network representations do not contain information on the specific CDI words (e.g., “mommy” is on the CDI, not “mother,” but “mother” is in more network representations than “mommy”). Given that network growth models cannot generalize to words outside of the network representation, we use the baseline age-specific CDI norms to compute...
3.1. Growth Modeling Equations. We consider three different network growth models originally proposed by Hills and colleagues [2] updated for our modeling framework. We formalize the growth models to provide mathematical equations and understanding as to how these models are distinct in the underlying growth mechanisms and in their ability to account for future lexical acquisition of individual children (see Figure 1 for a visual representation of predictions resulting from these models). For the mathematical formalization, each network graph can be represented as an adjacency matrix $N$ of size $|V|$, where $V$ includes all unknown and known words of the child that are part of the CDI vocabulary checklist and network representation. $N_{i,j} \in [0, 1]$ indicates the existence and weight of an edge from $i$ to $j$. We assume network representations are converted to binary such that $A_{i,j} \in \{0, 1\}$ (see below for motivation and more information). The threshold $\theta$, for $A_{i,j} = N_{i,j} > \theta$, is learned specifically for each model we consider such that the growth model is maximally predictive on the training data. We also define the binary network induced by the set of words a child can produce at a particular point in time as $K$, $c_i^M$ is the “importance” of node $j$, given network $M$. We presume edges of a network are defined by the chosen network definition. Edges form the basis for the calculation of a word’s importance. We additionally assume that if a word is learned, the word and all respective edges to known words are added to the child’s lexical graph. We do not consider words to be learned sequentially but instead predict learning of the joint set of words. This nonsequential assumption of our model is important as some growth processes assume that words are learned conditioned on the current vocabulary of the child, and thus the model estimates will differ based on the assumption of batch versus sequential learning. We choose batch learning as the data used to fit the model are themselves a set of words the child learns, and thus we avoid inference over the order in which these words are learned. We discuss this in more detail below.

All models make predictions only for unknown words. We define each network growth model as the function that maps $c_i^M$, the measure of the importance of word $i$, to $\delta_i$, the probability of learning word $i$, for each unknown word $i$. For some models, $M$ is based on the specific words known by child $x$, in which case we add an index $x$ to $\delta$, and to $M$ to indicate the role of that particular child’s language knowledge. This $\delta$ value, or node “importance,” simply provides a means to quantify the model’s estimate of the utility that an individual (unknown) word has if learned. This is likely different for each specific child and is likely influenced by the words the child knows. In our work here, we consider different types of centralities and their relation to the lexical graph to operationalize the importance of individual words. Other types of word-specific measures, such as frequency or word length, could also be used.

**Normative Age of Acquisition** refers to our baseline estimate of learning approximated by the child’s age, known vocabulary, and normative CDI reports. The normative CDI reports include, for each specific word, the percentage of children at a given age (rounded to the nearest month) that were reported to produce that word. We can then compute a $\delta_{cdi(x)}$ value for each unknown word. This is computed by considering the child’s age and linearly interpolating the normative CDI reports to calculate the percent of children at that given age who would produce the word.

Defining $B_{(i,m)}$ as the linearly interpolated proportion of children who know a particular word $i$ when the child is $m$ months old, we can compute the base rate probability of child $x$ learning all unknown words on the CDI ($\delta_{cdi(x)}$) for each word $i$

$$
\delta_{cdi(x)} = \frac{B_{(i,m)} - \theta}{\sum_{i=0}^{n} B_{(i,m)}}
$$

(2)
where \( m_x \) is the age in months of child \( x \) at the point of prediction.

In all of our network growth models, we include this \( \delta_{x,i,l} \) value. The performance of a model that only includes the \( \delta_{x,i,l} \) value in equation (1) is then our baseline model. This baseline prediction can also be used for all words not included in the network representation to allow comparison across network representations that include different words.

**Preferential Growth** assumes that words are more likely to be learned if they connect to nodes that are themselves well connected in the graph. A word is added to the graph, or learned by the child, in proportion to the total importance of each known word it attaches to. For example, looking at Figure 1, the word “car” is connected in the child’s known vocabulary \( K_{x,s} \), and thus this model assumes that words related to “car” are most likely to be learned next. Under this definition, the preferential growth model can be defined as

\[
\delta_{i,x} \propto \sum_{j \in K_{x,s}} A_{i,j} K_{j,s}.
\]

In this equation, we sum up the centrality \( c_j^{K_{x,s}} \) of known words only if there is a connection between the unknown word \( i \) and \( j \). We do this by using \( A_{i,j} \) as an indicator to ensure that word \( i \) connects to word \( j \). If word \( i \) is connected to word \( j \), we consider word \( j \) as contributing some increase in the probability of learning word \( i \). We define this increase in likelihood of learning as \( c_j^{K_{x,s}} \) (the centrality measures of word \( j \) as computed from the network of words known by the child \( K_{x,s} \)). We do this for all known words. This sum is then considered proportional to the growth value of word \( i \). For example, if we assume centrality is in-degree, this results in the sum of the in-degrees of the words \( j \) that \( i \) would connect to if learned. Note that the network of known words \( K \) is conditioned on the particular child’s vocabulary at the time of prediction and thus includes the subscript for child id \( x \).

This type of preferential growth mechanism would suggest that words are learned if they connect to highly important or central already known words. One possible cognitive mechanism that could drive this model is semantic or phonological differentiation, in which words are learned if they are similar to already known concepts (semantics) or sounds (phonology) [1]. One key feature of this model is that predictions of word learning is driven only by the connectivity of words that the child currently produces \( K_{x,s} \), with no influence of the global language environment.

**Preferential Acquisition** instead implies that words are learned based on their connectivity in the larger language environment, summarized in a network context as the “full language graph.” We consider this full language graph to be the graph constructed presuming the child knows all the words in our vocabulary assessment. Mathematically, the growth value of each word under preferential acquisition is defined as follows:

\[
\delta_i \propto \epsilon_{i}^{A},
\]

where \( A \) is the full binary network. This model relies on the idea that the more important a word is in the global environment, the earlier it is learned. Additionally, this model assumes that the full language graph approximates the *language environment* and *linguistic context* that is important to child language learning. What makes a word central varies on the specific network definition and centrality measure used to approximate the importance of individual words. For example, if we use in-degree centrality, this model assumes that words that have the most neighbors in the full language graph are those that are most likely to be learned next. Such a growth process might indicate that children are learning words that are contextually diverse (e.g., appear in a variety of context) and those words that are most likely to be learned earliest. This could be due to the fact that children can more easily learn word-object mappings when the object appears in many different environments [21], a fact that is naturally captured in the degree of a node in networks like free association graphs [2, 15]. Note that this model is not influenced by the child’s vocabulary graph, and thus there is no need to have an index \( x \) for the child. Individual-level differences simply emerge from the fact that children know different words, and thus normalization will result in different probability distributions for individual children.

![Figure 1: The three growth models with vocabulary predictions given a specific child’s vocabulary. Red words are words already known by the child, green words are words learned by the child, blue words are the 10 most likely wordsto be learned according to model predictions, and the yellow word (water) is a word that was correctly predicted as learned by a specific model. The size of the blue dots indicates the certainty the model gives to learning that specific word.](image-url)
Lure of the Associates bridges the gap between a model based only on the connectivity of words in the child’s known vocabulary and a model based only on the connectivity of words in the language environment. Here, a word is learned proportionally to node “importance” in the graph but conditioned on the child knowing at least one of the words that gives rise to the edge in the graph. We formalize this by defining the probability of learning word $i$ as the centrality of that word if that word were added to the graph (indicated here as the union of the known words of a particular child $K_i$ and word $i$). Words are learned if they are more central to the known vocabulary graph than other unknown words. For example, if a child’s vocabulary network has an animal component and a water component, bridging words like duck and fish might have higher betweenness centrality when added to the graph than other unknown words. This model presumes that the words that are most likely to be learned are words that will become most important in the productive vocabulary graph (in comparison to other unknown words) once learned. We define word importance as

$$\delta_{i,x} \propto c_i^{(K_i \cup i)},$$

when lure of the associates is the presumed model by which the network representations of small children grow.

Allowing for word importance to be based only on pairwise relationships in which at least one element of the pair is known suggests that children may need context and understanding to ground learning and allows lure of the associates to have a stronger relationship to the child’s known vocabulary than preferential acquisition and a stronger relationship to the language environment than preferential growth.

For the current analysis, the way in which we define “word importance” is based on not only growth models and computation of $\delta$’s but how we define centrality or $c$ in these models. Centrality measures, which are embedded in our calculation of growth values, calculate the role of each node in the graph. While there are many different types of centrality capturing different types of node importance, we consider three types of centrality that we believe are cognitively relevant and may capture some meaningful aspects of language acquisition in young children. The first is in-degree centrality as put forth and considered on normative vocabulary snapshots by Hills and colleagues [2]. We also consider undirected betweenness and closeness centralities. Though these measures are correlated, there are differences in terms of interpretation when using these centrality measures. Degree centrality models presume that the number of neighbors a word has is relevant to making a prediction of future word learning. Degree centrality is considered a more local measure of centrality as a node’s degree centrality is based only on the node’s immediate neighborhood and not the global location of the word in the full lexical network. Betweenness centrality instead suggests that words are more likely to be learned if they provide new and/or shorter paths between currently known words. Closeness centrality suggests that vocabularies may grow based on minimizing distance between words, even as the network grows in size. Both betweenness and closeness centrality are considered to be more global measures of centrality as changes in the global graph structure will influence the node-level centrality measures. On average, we can expect betweenness centrality to change more drastically than closeness when we change the graph structure.

3.2. Longitudinal CDI Data. To evaluate our models, we need detailed data of the words that a child learns through the course of development. As mentioned above, one well-established way to measure and characterize toddlers’ lexicons is to use vocabulary checklists, such as the MacArthur-Bates Communicative Development Inventory (CDI) [17]. The CDI checklist, completed by parents, indicates whether or not the child produces each word of a fixed set of words. These parent-reported vocabulary measures have been shown to be effective in evaluating children’s communicative skills up to 30 months of age [22, 23]. The CDI: Words and Sentences Toddler Form is a checklist of approximately 700 early words that are typically produced by the at least half of children by 30 months of age.

Longitudinal CDI data from 83 monolingual toddlers (37 females) were collected as part of a 12-month study, conducted at the University of Colorado Boulder, Colunga Lab. Recruitment for the study was done in three phases and was biased toward recruiting children that were learning language at a slower rate than their peers (classically called late talkers). Language ability was evaluated based on CDI percentile. CDI percentile is an estimate of the number of children one could expect who would have a vocabulary equal in size or smaller than the observed child’s vocabulary when controlling for age and gender. Observed CDI percentiles, as computed based on CDI norming data, spanned all language learning levels with an average learning percentile of 37.3 at the first of 12 visits and 61.3 at the end. The mean age of children was 17.5 months (range 15.4–19.3) at the first visit. On average, we have 10.9 CDIs (minimum of 2 and maximum of 12) for each child. Altogether, we have a total of 908 CDI forms. Figure 2 represents the type of longitudinal data utilized for modeling. For modeling purposes, we consider the change in vocabulary, or the difference between two sequential CDIs from the same child, to be a vocabulary snapshot, with the first CDI being the initial CDI and the second being the prediction CDI. In total, we have 825 CDI vocabulary snapshots.

One goal of our modeling approach is to predict the individual words a child is likely to learn next. To this end, we model how the network of an individual language learner changes over time by predicting when nodes will enter the graph. Figure 3 visualizes a learning trajectory of a child via four network graphs that are changing over time based on the addition of nodes to the graph. Edges are based on the McRae feature norms [18]. Our research goal is to construct a network growth model that captures the evolving networks structure through accurately predicting the individual words the child is likely to learn next.

4. Experimental Validation

For training individualized network growth models, we utilize cross validation. Training data consist of 60% of the
### Table

<table>
<thead>
<tr>
<th>Age (mo)</th>
<th>Sex</th>
<th>Voc. size</th>
<th>Dog</th>
<th>House</th>
<th>Zoo</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.2</td>
<td>F</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>17.1</td>
<td>F</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18.9</td>
<td>F</td>
<td>132</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19.3</td>
<td>M</td>
<td>257</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20.5</td>
<td>M</td>
<td>345</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

#### Figure 2: Example longitudinal CDI data for two children. Our models are trying to predict when a word enters the graph or the grey cells.

#### Figure 3: Network representation of a child’s growing productive vocabulary.

children and a total of 484 snapshots. Validation and test sets each include 20% of children. All snapshots of a single child were included in only one group. Because of the variability in number of snapshots of an individual child, we verified that the size of the data (e.g., number of individual snapshots) also had similar proportions to the 60/20/20 split of children’s data. Note that models were evaluated not only on performance of unseen snapshots but also on unseen children.

Each training example consisted of a vocabulary snapshot or a paired set of CDI reports collected at approximately one month intervals. The initial CDI was used to construct the child’s known vocabulary network. The growth value was then calculated assuming a specific network representation, growth process, and centrality, conditioned on the child’s current productive vocabulary where relevant (see equation (3) through (5) for specifications on how the growth values were calculated for the three different growth processes). The growth value is then combined with a base rate learning value from the CDI norms; these inputs are then converted to a probability through a logistic transformation as discussed in equation (1), with binary threshold \( \theta \), scale, and intercept parameters optimized and validated on training and validation sets, respectively. The resulting probability indicates, for each word not known by the child at the initial CDI, the probability that the word will be learned by the next CDI for that child. Observations are approximately a month apart, but the time between snapshots varies slightly across observations due to difficulties of scheduling.

Each model has a total of four parameters which we review in detail here. A threshold, converting the weighted graph to a binary graph, is optimized. We threshold the graph because the binary representation aids in interpretability of results and provides quicker convergence and more stability to the models. We note that this is a simplifying assumption made out of computational convenience. Although the acquisition of a word is generally a protracted process, with children remembering, forgetting, and refining the referential scope of a word over a period of time [24, 25], the data with which we have to build these networks are a binary judgment on the part of the parents (whether the child says or does not say the word), and we have no direct access to the child’s possible specific representations of the word. In addition, as we are exploring different lexical representations and attempting to learn something from comparing across them, optimizing the threshold for each representation seems like a fair thing to do. Finally, the use of a binary network makes structural comparisons across networks more interpretable. The threshold can be thought of as a means to cancel out distributional and measurement noise and to instead highlight connections and relationships that are strong enough to garner attention of a child rather than a weighted network which would provide a notion of the importance of relationships between objects. With this binary graph and the child’s known vocabulary at the initial CDI of our snapshot, we then compute a baseline probability \( \delta_{\text{cdi}(\text{init})} \) and the delta value based on the network growth model as defined in Section 3.1. We then fit a logistic regression, converting these growth values and CDI baseline values to probabilities. Note that we are fitting both the network growth value and the growth value from the CDI age of acquisition norms simultaneously and for each network representation. This is used to account for the fact that some words are learned based on developmental rather than linguistic trajectories. We repeat our optimization procedure using expectation maximization on the threshold to convert the weighted graph to a binary graph.
probabilistically selecting a threshold and learning the optimal logistic regression parameters and network threshold based on optimal negative log-likelihood values on a validation set. Because the network size varies based on the network representation, we assume words on the CDI, but not in the network representation, are learned according to the normative, age-specific acquisition baseline model, e.g., equation (1) with $\beta_1$ fixed to 0.

Model selection is based on minimizing negative log-likelihoods for predictions to unknown words in the validation set. This measure penalizes both overestimation and underestimation of learning specific words such that if the model is highly confident a word is not learned and the word is learned; this miscalculation contributes equally to the error of confidence that a word is learned when it is not. We also note that, because we are only including predictions to words that are unknown at the initial CDI in the snapshot, this measure overrepresents children who have small vocabularies because they have many more CDI words that they could possibly learn. Using the validation data, we compare model performance to the age of acquisition baseline model. Because we are exploring network representation, network growth mechanisms, and centrality, we select a subset of the most predictive models to extend to the test set and to discuss in terms of insights into lexical acquisition.

5. Results and Discussion from Network Growth Framework

5.1. Macro Level: Effect of Network Structure. We begin by characterizing the structure of the network representations on the maximal lexicon. To create the network representation, we take the overlapping words between network representation and the CDI norms. We consider words and their variants, like shoe and shoes, to be equivalent and, when necessary, average their representations. Alignment between CDI words and network representations results in networks of different sizes ranging from 133 words to the full 677 words included in our longitudinal CDI assessment. In Table 1, we compare the structure of the observed networks to two types of random models. We note that the original network representations are weighted. We use the threshold as learned by the best performing growth model and centrality when characterizing the network structure here (see Table 2 for the threshold values). We compare the structure of the lexicon to two random models to better understand how representation of these language graphs differs from randomly generated semantic graphs.

The first comparison of the observed language graph to randomly generated, size-matched networks is the configuration model. The configuration model results in the construction of random network variations where the degree distribution matches the degree distribution of the observed network. We abbreviate this model in Table 1 as CM. We also generate a random model via a variant of preferential attachment [8] as proposed in the language acquisition paper of Steyvers and Tenenbaum [1]. We abbreviate this model as ST. In the Steyvers and Tenenbaum model, the network starts with a few seed words, and then at each iteration, an attachment word is selected proportional to the word’s (current) degree. Once an attachment point is selected, a new node is added, with an edge between it and the attachment point. Then, the new node is connected to neighbors of the attachment point with some probability $\alpha = 0.9$. Neighbors are sampled until the new node has a degree equal to the mean degree or until every node in the current vocabulary is considered (see [1] for more information on this model). Because of the iterative edge building of the ST model, this model does not always converge on dense graphs since there are not enough available neighbors to maintain the high observed degree distribution at early iterations; thus, we do not include the results for the ST model on phonology as the resulting random network deviates greatly from the observed network even on network density. The lower density of the ST model comes from the fact that the observed graph has a density near 1 and edges added early cannot maintain the overall network density of the observed phonological network. We present the size of the graph, the density, the average degree, transitivity (e.g., the probability that a is connected to c given a is connected to b and b is connected to c), mean geodesic, graph diameter or maximal shortest path, and the assortativity coefficient for (1) our observed networks, (2) configuration models (CM), and (3) Steyvers and Tenenbaum preferential attachment models (ST). We average over 100 runs of each random model in Table 1 and report mean estimates as well as standard deviations around these estimates.

When comparing the observed early semantic graphs to those generated by the random variants, we find that the local structure and assortativity of the observed networks cannot be well captured by the random graphs we considered. The phonological network representation is difficult to compare since the average node degree is almost the size of the full graph, and thus we focus instead on the semantic networks. We find that our observed networks have more local structure as captured by transitivity than would be expected by the degree distribution alone (in comparison to CM). Looking at the ST model, we find that the process of adding edges results in a slightly less dense network as nodes added early cannot form edges with enough neighbors. The transitivity measure of the networks generated by this model is close to the observed value in our semantic networks due to the fact that edges are added with high probability and between direct neighbors of the attachment node. Both random models, however, fail to recreate the geodesic distance of these observed networks, underestimating the distance. Both models also fail to recreate the assortativity observed in the language networks, and in two cases, the random models reverse the direction of this assortativity.

Although Table 1 aims to summarize the network representation as a whole, we use an individual child’s specific productive vocabulary to induce a network which our growth model uses to predict learning of unknown words. To better understand the structure of these lexical networks, we consider the specific network structure of our snapshot data. In Figure 4, we plot the density, average local clustering
as assortativity coefficients. It is easy to see that in most cases, the configuration model cannot capture transitivity or the local structure and no random model can account for the observed mismatch, and a preferential attachment algorithm applied to early acquisition (ST) as discussed in Steyvers and Tenenbaum’s 2005 paper [1]. We also include all CDI snapshots in this figure to highlight networks, this measure does not vary in relation to developmental trajectory. Understanding what might be contributing to children learning language at a slower pace than their peers is complicated by the fact that small vocabularies have relatively high variability when measured with CDI checklists as there are more possible words that could be learned.

However, looking at network representations of these snapshots of children with small vocabularies and children with low CDI percentiles, we see stark differences in network structure, providing a promising direction for future work looking at diagnostics and interventions on late talkers. When comparing the best fit smoothed average (local polynomial regression fitting) of the left most data across the top and bottom rows, we find that children who have lower percentiles have larger giant components than younger children with similar vocabulary size in the case of the Nelson and McRae network representation, but that this giant component size becomes connected to all words in the network later in development. In the case of the Nelson network, we find that local structure as measured by transitivity is much higher in late talkers than in children with small vocabularies, even though late talkers often have small vocabularies. We also find that density and transitivity are higher for late talkers than younger children in the phonological network representation as well. These differences could suggest that late talkers and/or younger children may learn differently and will be a focus of our future research.

Table 1: Network summary statistics of network size (included next to the observed network name and consistent across all random models), network density, mean degree (\(\langle k \rangle \)), measure of transitivity (trans.), mean geodesic distance (geodist.), diameter (diam.), and assortativity coefficients (assort.).

<table>
<thead>
<tr>
<th>Network</th>
<th>Density</th>
<th>(\langle k \rangle )</th>
<th>Trans.</th>
<th>Geodist.</th>
<th>Diam.</th>
<th>Assort.</th>
</tr>
</thead>
<tbody>
<tr>
<td>McRae (133)</td>
<td>0.191</td>
<td>25.33</td>
<td>0.524</td>
<td>1.99</td>
<td>4</td>
<td>-0.037</td>
</tr>
<tr>
<td>McRae CM</td>
<td>0.191</td>
<td>25.33 (0.36)</td>
<td>0.309 (5e-3)</td>
<td>1.86 (0.01)</td>
<td>4.1 (0.31)</td>
<td>0.016 (0.017)</td>
</tr>
<tr>
<td>McRae ST</td>
<td>0.187</td>
<td>24.81 (0.02)</td>
<td>0.532 (7e-3)</td>
<td>1.85 (0.01)</td>
<td>3.9 (0.30)</td>
<td>-0.142 (0.011)</td>
</tr>
<tr>
<td>Nelson (545)</td>
<td>9.8e-3</td>
<td>5.38</td>
<td>0.153</td>
<td>4.65</td>
<td>12</td>
<td>0.012</td>
</tr>
<tr>
<td>Nelson CM</td>
<td>9.7e-3</td>
<td>5.36 (0.07)</td>
<td>0.027 (5e-3)</td>
<td>4.04 (0.04)</td>
<td>9.8 (0.78)</td>
<td>-0.004 (0.018)</td>
</tr>
<tr>
<td>Nelson ST</td>
<td>9.1e-3</td>
<td>4.99 (0.01)</td>
<td>0.109 (6e-3)</td>
<td>3.07 (0.05)</td>
<td>6.5 (0.31)</td>
<td>-0.137 (0.013)</td>
</tr>
<tr>
<td>Phono (677)</td>
<td>0.956</td>
<td>646.34</td>
<td>0.987</td>
<td>1.04</td>
<td>3</td>
<td>0.013</td>
</tr>
<tr>
<td>Phono CM</td>
<td>0.952</td>
<td>643.67 (0.12)</td>
<td>0.983 (1e-4)</td>
<td>1.04 (1e-4)</td>
<td>3 (0)</td>
<td>-0.004 (7e-4)</td>
</tr>
</tbody>
</table>

Note. We consider our observed networks, as well as networks constructed via configuration modeling (CM) networks (in which the degree distribution of the observed network is matched), and a preferential attachment algorithm applied to early acquisition (ST) as discussed in Steyvers and Tenenbaum’s 2005 paper [1]. It is easy to see that in most cases, the configuration model cannot capture transitivity or the local structure and no random model can account for the assortativity coefficients.

Table 2: Best performing models on each of the network representations.

<table>
<thead>
<tr>
<th>Network</th>
<th>Model</th>
<th>Cent</th>
<th>Thresh</th>
<th>Density</th>
<th>llk</th>
<th>CDI llk</th>
<th>(\beta_{cd})</th>
<th>(\beta_{\delta})</th>
</tr>
</thead>
<tbody>
<tr>
<td>McRae</td>
<td>Lure</td>
<td>Close</td>
<td>0.091</td>
<td>0.192</td>
<td>0.402</td>
<td>0.432</td>
<td>0.143</td>
<td>0.759</td>
</tr>
<tr>
<td>Nelson</td>
<td>Lure</td>
<td>Degree</td>
<td>0.011</td>
<td>0.010</td>
<td>0.384</td>
<td>0.394</td>
<td>0.651</td>
<td>0.357</td>
</tr>
<tr>
<td>Phono</td>
<td>Lure</td>
<td>Close</td>
<td>0.101</td>
<td>0.956</td>
<td>0.364</td>
<td>0.380</td>
<td>0.377</td>
<td>0.672</td>
</tr>
</tbody>
</table>

Note. All models reliably outperform random (CDI llk) when applied to validation data. The network threshold (thress) and network density as well as the log-likelihood of the model and the baseline CDI model are reported. The normalized logistic transform of the network (\(\beta_{\delta}\)) and CDI norms (\(\beta_{cd}\)) is reported for comparison of the importance of the CDI norms versus the network word features.
utility. In total, we consider 27 models (3 network representations by 3 growth models by 3 centrality models) in addition to a model that includes only the delta values from the CDI norms which assumes that the network growth model contributes no information to the model. Because the network representations include a different number of words and we would expect better performance of our network models on words that are in the graph, there is a unique CDI norm baseline for each representation. We can also compare across representations by using the CDI norm baseline for words that are not in the specific network graph. We find this type of analysis masks much of the impact of the network representations especially because the McRae network has only 133 words. We thus neglect this comparison for the remainder of the paper.

We begin by comparing performance of all our models to the word-matched CDI model. Figure 5 plots the improvement of model performance for each network representation aggregated by growth mechanism (Figure 5(a)) and centrality measures (Figure 5(b)); zero indicates performance of the baseline CDI model. The y-axis indicates improvement in percent of log-likelihood error over our baseline model. The x-axis is organized by network representation, but the specific x-values within a network representation are not meaningful and are only used for readability. Positive values in Figure 5 indicate the model outperforms the network-specific baseline model. Note that a specific model can perform worse than, or near to, the baseline because here we are considering performance on the validation data. Performance near zero may suggest overfitting or high parameter sensitivity.

One clear result that can be seen from this figure is that there is not a single dominating growth mechanism or centrality. If this were the case across all network representations, we would expect clear line separation of the models in terms of growth mechanism or centrality and a similar color gradient across the three network representations. We in fact only see clear separation in the case of the Nelson model where Lure of the Associates is the only model that shows a reliable improvement over all other models we consider (middle left panel). Even considering the Nelson representation, we still see a substantial interaction of centrality within the growth process of lure of the associates (middle right panel). The fact that the graphs are not easily separable along growth mechanism (Figure 5(a)) or centrality (Figure 5(b)) strongly suggests an important interaction between the mezzo (growth process) and micro (node-level interaction) levels of analysis in the domain of modeling acquisition.
In the three network representations we have chosen for our current analysis, all models perform at or above chance. However, Figure 5 suggests that there is no clear main effect of growth mechanism or centrality that is consistent across all network representations, or even within a network representation. Instead of evaluating the significance of each model, we choose to focus only on the best model for each network representation. We summarize these models in Table 2. We report performance of the model in terms of negative log-likelihood and include the performance of the CDI baseline model on the specific subset of words included in each network representation as well such that the reader can understand the improvement in terms of average negative log-likelihood of the network model as compared to the baseline CDI only model. Here we find that \textit{Lure of the Associates} is the best performing growth mechanism regardless of network representation. The fact that \textit{Lure of the Associates} is the best growth process supports the idea that, when predicting future acquisition, we should consider both the child’s existing lexical knowledge as well as the structure of language in the child’s environment.

In considering the mezzo level, beyond finding that \textit{Lure of the Associates} is the best performing model, rarely do we see \textit{Preferential Acquisition} significantly outperforming the baseline model, suggesting that it is not a useful mechanism for predicting the vocabulary growth of an individual child. Recall that \textit{Preferential Acquisition} assumes that words are learned proportionally to their centrality in the full vocabulary graph. Previous results found this mechanism (using degree centrality) to be the most accurate model when accounting for normative acquisition trends [2] (we replicate their results using our models on normative acquisition, confirming their finding that preferential acquisition is the best performing model for normative acquisition). The failure here to account for individual vocabulary growth is likely because this model does not adapt well to individual differences and, in fact, does not adjust predictions to a child’s vocabulary knowledge in any meaningful way. The inability of this model to predict individual acquisition trajectories and the success of \textit{Lure of the Associates} suggests two main findings: (1) normative acquisition is quite different than the acquisition of any particular child and (2) the content of the child’s vocabulary is important and predictive of which words a child is likely to learn next. \textit{The Lure of the Associates} increased performance over preferential growth may also suggest that children need grounding in their productive vocabulary for connectivity in the lexical network to aid in learning. This poses an interesting direction for future work on interventions related to accelerating vocabulary acquisition.

5.3. Micro Level: Effect of Centrality. Analyzing the role of network centrality suggests that global centrality measures (e.g., determined and influenced by the structure of the full graph) are as accurate as local measures (e.g., determined only by local neighbors). In the case of global measures, the addition of a node may change the full centrality of the network much more drastically than in the case of the local centrality measures, but this potentially large change does not seem to aid the predictive power of these growth models. The most local centrality measure we consider is \textit{in-degree} as this measure only considers immediate neighbors whereas the more global measures are closeness and \textit{betweenness} as these measures include contributions from all nodes in the graph. Global centrality measures particularly affect \textit{Lure of
the Associates, as this model selects words for learning that most substantially increase the connectivity of the known vocabulary structure. While we see that closeness centrality is the most predictive in two out of the three growth processes we consider, we do not see a clear effect of centrality. It seems there is a complex interaction between growth model and centrality that requires further investigation in future work. Another possibility is that the performance of these models could interact with aspects of the learner such that a particular growth process or centrality is most predictive for certain types of learners. We explore this possibility below.

Even while our results do not strongly indicate a specific and clear effect of centrality, our results confirm that the chosen definition at each layer of analysis (micro, mezzo, and macro) does affect predictive performance. We additionally see that the growth processes have a substantial effect on our ability to generalize to unseen language trajectories and that, for the purposes of individual modeling, certain models are not flexible enough to predict the lexical acquisition of individual children above the predictive accuracy of CDI baseline. Finally, we show evidence for microeffects emerging from the way we choose to operationalize node importance. Taken collectively, these results indicate that consideration of the different levels of analysis—and importantly, of the interactions between levels—can help us capture and model the full complexity of the overall process of acquisition, while providing potentially novel hypotheses that can be tested further in modeling work as well as empirical studies on young children.

6. Predictive Accuracy on Unseen Children

For evaluation on the test data, we consider only the models in Table 2 to avoid statistically biasing our sample because these models were most predictive when generalizing from training to validation data. Looking at Table 3, and specifically the coefficients of the models, we can see that the standardized (the input features have been standardized such that the features have the same scale and standard deviation; this allows one to directly compare β coefficients) β coefficients suggest that there is a significant contribution of the network-based node importance (βδ) as well as the CDI-based δ values (βδCDI). We can also see that the mean likelihood scores across all models are lower when including the network-based δ measures. Note we did not perform any statistical verification that these models reliably outperformed the CDI baseline at this time as to not bias our assessment on the test dataset. When extending our models to the test set (the results of which can be seen in Table 3), we explore different types of questions than those questions used for model selection and perform statistical analysis to verify if the network representation improves prediction accuracy.

In model selection, we optimized over four different parameters with respect to the average negative log-likelihood over all predictions to unseen words. We use the validation data to evaluate parameter fit. We considered the network threshold (to convert the network representation from a weighted network to a binary network), the growth process, the network centrality for each network representation, and the β weights for the CDI and network-based growth values. Because the binary network representation, and thus the calculations of the δs for the network representation, is affected by the threshold, we optimize the network threshold independently for each growth process and centrality measure. Note that minimizing log-likelihood is a slightly skewed measure because we are only predicting the learning of words (e.g., if a word is already known by the child, we do not make a prediction as to whether the word stays learned). This means that when minimizing log-likelihood over the whole dataset, younger children and children who know fewer words are overly represented in this sample. Additionally, it can easily be argued that we should care more about words that are learned than about words that stay unknown. This discrepancy between model optimization and research questions provides an interesting additional evaluation to our selected models.

We thus introduce additional measures in order to evaluate our selected models. The first measure is the average negative log-likelihood across unseen snapshots and has a similar interpretation to the log-likelihood values previously reported. Using the trained models, we make predictions individually to each unseen snapshot and compute the log-likelihood on that snapshot. We then average across snapshots. We can then easily compare, via paired t-test, whether our specific model outperforms the CDI baseline model. We additionally compute the percent overlap (k ∩ k′). In this calculation, we assume that we know how many words (k) a child learned from one snapshot to the next and then compute the percentage of those words that are in the top k′ words as predicted by the model. Additionally, we include the area under the curve (AUC) of the receiver operating characteristic (ROC) measure that considers the false-positive rate and the true-positive rate as the threshold for converting probabilities from 0 to 1 changes. The AUC summarizes this plot by computing the area under the curve constructed via varying the threshold. Also included are accuracy, precision, and recall values, computed at the population level assuming the best threshold from the ROC calculation. The results can be seen in Table 3.

The results included in Table 3 show that our network growth models are capable of predicting future word acquisition of individual children with higher accuracy than our baseline models (as all models are significant under multiple comparison correction with α = 0.05). We also see that the nature of the network representation can affect our ability to improve over the baseline model as indicated by the fact that some representations such as McRae result in a lower accuracy than the baseline CDI-only model. Similarly, the Nelson network model has a lower percent overlap than the CDI baseline model. This suggests that some models may be better at predicting acquisition of words (e.g., improvement in percent overlap), while other models may be better at predicting what words are least likely to be learned as well as those words that are learned (e.g., AUC or llk measures). These are important directions for future exploration as it is unclear whether teaching children words that they are about to learn, or words that they are unlikely to
learn, or something in between, is the best means of intervention for long-term changes in the rate at which children learn words.

We also find strong evidence for an influence of the child’s current vocabulary on the accuracy of our predictions as captured by the improved performance of the Lure of the Associates over the model that uses only information related to normative acquisition rates for words. These results suggest that modeling language acquisition benefits from the inclusion of interactions among the specific words a child knows. This finding highlights the importance of looking at the developing lexicon as a system, one that is likely to follow different developmental paths in different individuals. Additionally, we take these results as a promising validation of the cognitive insight and predictive modeling capabilities of a united network modeling framework that incorporates macro, mezzo, and micro levels of analysis.

7. Modeling Development

We are interested in a developmental perspective, beyond simply predicting future language learning. It is possible that these models are more accurate during certain periods in development or for earlier/later learned words. We note that this analysis is post hoc and that models were neither trained nor optimized to capture developmental effects. To consider these developmental effects, we take a naive approach of ordering the predictions by specific features we believe might be relevant for learning. Because each model has an individual baseline model optimized to predict the lexical items in a particular network representation, we compute the difference between the baseline CDI model for each representation and the network predictions such that values greater than zero indicate that the network model outperforms the baseline model for that specific snapshot. Although we showed above that all network models statistically outperform the baseline models, this does not mean that there is a performance boost for all children equally across development. Thus, we aggregate the individual likelihood predictions across a theoretical ordering of when words are learned (age of acquisition effects, e.g., certain words are on average learned earlier) or across snapshots (for developmental effects) to uncover trends in predictive accuracy of our network models.

Plotted in Figure 6, we show performance differences as compared to the CDI age of acquisition baseline. We construct orders based on the (1) average age a word is learned (AoA, average age of acquisition) as calculated based on normative acquisition trends (Figure 6(a)), (2) age of the child at time of prediction (Figure 6(b)), (3) CDI percentile of the child at the initial CDI of the snapshot (Figure 6(c)), and (4) vocabulary size of the child when the first CDI is collected (Figure 6(d)). Plotted points are the differences between the CDI baseline and the network model, colored based on network representation. We then fit a local polynomial regression which plots the smoothed locally weighted averages as a line to indicate performance of the models with respect to certain features of the child or vocabulary that we find interesting. We first note that when considering the word level, there are no clear trends as to when the network model outperforms the baseline model. However, when we organize the data based on child features, such as the age of the child at prediction, the smoothed curve of the log-likelihood difference suggests that certain representations perform better than others at different periods of development. Because age, percentile, and vocabulary size are all intercorrelated, we would expect the fitted lines to show similar trends. We find that the CDI age of acquisition baseline model outperforms our network models for some points in development; however, our network models are still significantly outperforming their corresponding CDI baseline models (see Table 3). We note that these results are only suggestive as there is a lot of noise in the data and this is a post hoc analysis. However, it suggests interesting trends that may be worthy of further investigation.

Performance indicates that the phonological network representation outperforms the baseline model, particularly early in development and toward the end of development. The fact that the phonological representation is most useful early in development also aligns with intuition that early in language learning, a child is limited by which words they can articulate intelligibly as much, if not more, than which words they know the meaning of. The improvement in predictive accuracy toward the end of development will take further investigation and more targeted theories as to the role of phonology later in the acquisition process. Another period of development in which our network models outperform baseline is for older children, children with high percentiles, and children with large vocabulary size. If a child knows nearly all the words, the network model can use information about the structure of those unknown words to pick up on statistically meaningful relationships, whereas the CDI

Table 3: Evaluation of model performance on test data.

<table>
<thead>
<tr>
<th>Model</th>
<th>V</th>
<th>Snap. llk</th>
<th>p value</th>
<th>k ∩ k'</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>McRae</td>
<td>133</td>
<td>0.478</td>
<td>0.018</td>
<td>34.1</td>
<td>0.689</td>
<td>0.609</td>
<td>0.225</td>
<td>0.702</td>
</tr>
<tr>
<td>M. CDI</td>
<td>133</td>
<td>0.516</td>
<td>0.030</td>
<td>37.9</td>
<td>0.761</td>
<td>0.677</td>
<td>0.256</td>
<td>0.726</td>
</tr>
<tr>
<td>Nelson</td>
<td>534</td>
<td>0.531</td>
<td>0.018</td>
<td>38.5</td>
<td>0.764</td>
<td>0.691</td>
<td>0.270</td>
<td>0.668</td>
</tr>
<tr>
<td>N. CDI</td>
<td>534</td>
<td>0.561</td>
<td>0.008</td>
<td>34.6</td>
<td>0.734</td>
<td>0.690</td>
<td>0.238</td>
<td>0.627</td>
</tr>
<tr>
<td>Phono</td>
<td>677</td>
<td>0.460</td>
<td>0.002</td>
<td>31.6</td>
<td>0.767</td>
<td>0.697</td>
<td>0.207</td>
<td>0.664</td>
</tr>
<tr>
<td>P. CDI</td>
<td>677</td>
<td>0.502</td>
<td>0.001</td>
<td>34.1</td>
<td>0.689</td>
<td>0.609</td>
<td>0.225</td>
<td>0.702</td>
</tr>
</tbody>
</table>

Note: We include the size of the network representation (or number of words the model makes predictions for), the average negative log-likelihood for each unseen snapshot. We also report the p value of an F-test for nested models comparing the CDI-only model to the model containing word importance based on the network growth process. We include the overlap between the words learned by the child and the top words as predicted by the model or percent overlap (k ∩ k'). Considering all predictions, we report the area under the curve (AUC) using ROC, accuracy, precision, and recall.
model must rely on global and aggregated trends across all learners, neglecting individual differences. This improvement for children who know more words is further evidence that the child’s current vocabulary aids in future language learning and that our network models may be capturing this relationship in meaningful and predictive ways.

These results collectively support the idea that individual modeling may be useful in capturing learning and developmental trends specifically for children who are far from the sample average. We can leverage this to our advantage particularly because the population we are most interested in modeling, namely, children with a low CDI percentile, is better predicted by these network models than the CDI baseline model. We find that the phonological network models are better than the CDI baseline norms particularly for very small CDI vocabularies and for children who have low CDI percentiles. This, however, would not explain how the phonological network is particularly predictive as compared to the baseline in the children with small vocabularies and low CDI percentiles. It is possible that our phonological network model may be capturing something about the nature of the deficit underlying the delay in some of these children—that being able to produce specific combinations of sounds is at least part of what is holding back their language development and thus offers a possible avenue for intervention [27, 28].

The high amount of noise in our plots when we aggregate by child features could be because there is a lot of variance in the network’s explainability or because we have yet to find the features of a child’s lexicon or child’s development that correlate with the range when our network representations are most useful in predicting future language learning. In future work, we aim to further investigate not only network growth models that are capable of modeling individual differences but also hope to gain a better understanding of when our network models will be useful versus when the baseline CDI model is equally reliable.

8. Discussion and Future Direction

The network-based approach to modeling acquisition provides a unifying framework for studying the complex process
of word learning, allowing researchers to investigate individual differences and predict future vocabulary growth of individual children. We find that the definition of the edges in a network, the assumed process of network growth, and the network measure chosen to operationalize importance of words dramatically affect our ability to predict future acquisition. The importance of the three levels of analysis can be seen in the varying fits across our combinations of models. Although we find evidence, at particular points in development, that one network representation is clearly more accurate at predicting lexical acquisition of young children, this is an area requiring future study—if we understand contexts in which these network models are most useful for prediction, we will be able to use these models to provide insight into diagnostics and interventions.

Taken collectively, interesting results emerge that may provide future directions of study and possible interventions. For example, the phonological edit distance (Phono.) network captures acquisition trajectories of younger children and children with lower percentiles. This is in line with findings that suggest that late talkers are disproportionately behind in their production skills (the words they say) rather than their comprehension skills (the words they understand) [29]. We also find that at certain points in development, the CDI baseline model is comparable in predictive accuracy to our network-based approach. This suggests the possibility that there are attentional changes during the course of learning and, potentially, that later talkers or younger children learn differently than their peers. Again, there is empirical evidence from behavioral tests with toddlers to suggest that this is the case [30–32] This type of network modeling framework may allow for us to not only model differences in these groups but to explain the process of acquisition that leads to these differences.

The results suggest that phonological and semantic features are both important and relevant to language learning. While phonological network structure varies greatly from semantic network structure and network model results have shown differences in acquisition in relation to phonological networks [20, 33, 34], both of these individual network representations are useful in predictive models of lexical acquisition. In the future, we hope to jointly consider the effects of these levels of analysis in predictive modeling and as a means to understand the process of language development, possibly by building ensemble models within this framework or by extending this approach to multiplex representations [35, 36].

Our results also challenge previous work in the domain of network-based approaches to modeling acquisition. In the case of normative language acquisition modeling, it has been shown that Preferential Acquisition outperforms models based on the child’s current vocabulary knowledge [2]. Here, however, we found that a model of preferential acquisition is unable to account for language growth at the level of individual children. All accurate predictive models of individual child trajectories use information about the child’s productive vocabulary. We found strong evidence that Lure of the Associates, a model that considers both information about the child’s current lexical knowledge and the language learning environment of the child as captured by the network of the full language graph, is the most predictive model regardless of the network representation we choose. This provides strong support to the notion that individual differences in learning can be related to both the language learning environment of the child and differences in the way that the child learns language. These results also highlight the need for modeling at the level of the individual rather than at the aggregate level such as via the age of acquisition norms. Although our models are ambiguous about how a child learns a specific word, the words themselves may have important and useful cues as to which are the relevant features that guide future lexical acquisition. These language learning cues may be related to the physical and linguistic environment that the child is immersed in or even to the child’s specific interests. Either way, our modeling results strongly suggest an important contribution of the known vocabulary on future vocabulary growth.

In the best fitting models for each network representation, we find an impact of centrality measures on accuracy of the network growth models. One interpretation of this finding is that both global structure and local information are important to the relationship between emerging language graph and future language learning. This interplay between local and global network structure may be especially important for correcting language learning delays such as those of late talking children. The role of global centrality measures (betweenness) and local ones (degree) suggests that instead of teaching just a few words to help get children back on track, we may need to alter the connectivity of the graph instead. For example, previous training studies [37, 38] have shown that teaching words for shape-based categories (like “spoon” or “ball”) accelerates vocabulary growth in typically developing children. Our results suggest that rather than having a static list of words that might be good for toddlers to learn, a goal may instead be to achieve a certain type of connectivity, for example, strengthening a nascent cluster of concepts (e.g., animal names), or teaching a specific word (e.g., “chicken”) that might serve to link two clusters or categories (e.g., animal and food word clusters). We also note that we only consider network centrality measures here, but there are other ways of quantifying a node’s importance. In the future, we hope to combine centrality measures with other non-network-related measures such as frequency or concreteness. While network connectivity seems to be useful in modeling language acquisition, other measures may provide additional support.

We note that this work is not the first work to use computational models to explore individual differences in development. In fact, there is a great deal of work focusing on capturing the structure of the child’s language environment with fewer simplifying assumptions of our preferential acquisition and lure of the associates model [39–41]. Much of these computational modeling results suggest that children may be learning distributional information directly from the environment and that learned distributional information can explain production and use [41, 42]. These models are an important part of the overall quest to understand the acquisition process and complement our work.

Complexity
by studying the language learning environment. Our work instead makes many simplifying assumptions about this learning environment but with the aim of capturing the learning process and the impact of the child’s current vocabulary knowledge on the overall process of acquisition.

The complexity of these network modeling results underscores the depth of the challenge that comes with modeling individual acquisition trends. Although here we present a first step in building up an accurate predictive network growth model, much more work is needed to explain why certain models perform in disparate ways, for different words, and for different language learners. In the future, with more data and more sophisticated network models, we hope to capture language learning with higher accuracy. Improving accuracy will also allow for the development and investigation of mechanistic models that offer explanations as to why certain children follow a specific language acquisition trajectory. We are particularly encouraged by the strong improvement of the network model over the CDI age of acquisition baseline model specifically for individual children who are learning language at a slower rate. Critically, although the accuracy in predicting future acquisition is a meaningful benchmark, one benefit of network analysis models over machine learning models is that the former models imply a possible mechanism of learning. Achieving a mechanistic understanding of the forces that shape early language acquisition is crucial in finding ways to improve the vocabulary of those children who are at higher risk of persistent language delays. In the future, we hope this work and other similar approaches of modeling at the level of the individual can pave the way for diagnostic and intervention tools capable not only of predicting but also explaining individual acquisition trends.

Data Availability

The data are not publicly available, please contact the first author for the code to run the computational models.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by an award from the John Merck Scholars Fund and by NICHD (grant R01 HD067315) to Eliana Colunga. The authors also thank Michael Mozer, Matt Jones, Aaron Clauset, Tamara Sumner, and Massimo Stella for their helpful discussion and ideas around this project.

References


