1   The growth of children's semantic and phonological networks: insight from 10 languages

2   Abdellah Fourtassi[1], Yuan Bian[2], & Michael C. Frank[1]

3   [1] Department of Psychology, Stanford University

4   [2] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

5   Author Note

6   Abdellah Fourtassi

7   Department of Psychology

8   Stanford University

9   50 Serra Mall

10   Jordan Hall, Building 420

11   Stanford, CA 94301

12   Correspondence concerning this article should be addressed to Abdellah Fourtassi,

13   Postal address. E-mail: afourtas@stanford.edu

Abstract

Children tend to produce words earlier when they are connected to a variety of other words along the phonological and semantic dimensions. Though these semantic and phonological connectivity effects have been extensively documented, little is known about their underlying developmental mechanism. One possibility is that learning is driven by lexical network growth where highly connected words in the child's early lexicon enable learning of similar words. Another possibility is that learning is driven by highly connected words in the external learning environment, instead of highly connected words in the early internal lexicon. The present study tests both scenarios systematically in both the phonological and semantic domains across 10 languages. We show that phonological and semantic connectivity in the learning environment drives growth in both production- and comprehension-based vocabularies, even controlling for word frequency and length. This pattern of findings suggests a word learning process where children harness their statistical learning abilities to detect and learn highly connected words in the learning environment.

Keywords: Word learning; semantic network; phonological network; network growth; cross-linguistic analysis.

The growth of children's semantic and phonological networks: insight from 10 languages

Introduction

What factors shape vocabulary learning over the course of early childhood? To investigate this question, scientists have adopted multiple research strategies, from conducting controlled laboratory experiments (e.g. Markman, 1990) to analyzing dense corpora capturing language learning in context (e.g., Roy, Frank, DeCamp, Miller, & Roy, 2015). One prominent strategy consists in documenting the timeline of words' acquisition and studying the properties that make words easy or hard to learn (e.g., Goodman, Dale, & Li, 2008; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). For example, Goodman et al. (2008) found that, within a lexical category (e.g., nouns), higher parental frequency is associated with earlier learning. Researchers have studied the role of several other factors such as word length and the mean length of utterances in which the word occurs (e.g., Braginsky, Yurovsky, Marchman, & Frank, 2019; Swingley & Humphrey, 2018).

Besides word-level properties, the structure of the lexicon (that is, how words relate to one another) is also linked to the Age of Acquisition (AoA) of words. The lexical structure can be characterized in terms of a network where each node represents a word in the vocabulary, and each link between two nodes represents a relationship between the corresponding pair of words (e.g., Collins & Loftus, 1975; Luce & Pisoni, 1998). Previous studies have investigated early vocabulary structure by constructing networks using a variety of word-word relations including shared semantic features (McRae, Cree, Seidenberg, & McNorgan, 2005), target-cue relationships in free association norms (Nelson, McEvoy, & Schreiber, 1998), co-occurrence in child-directed speech (MacWhinney, 2014), and phonological relatedness (Vitevitch, 2008). These studies have generally found that children tend to produce words that have higher neighborhood density (i.e., high connectivity in the network) earlier, both at the phonological and the semantic level (Carlson, Sonderegger, & Bane, 2014; Hills, Maouene, Riordan, & Smith, 2010; Hills, Maouene, Maouene, Sheya, &

Smith, 2009; Stella, Beckage, & Brede, 2017; Storkel, 2009).

While most studies have focused on the static properties of the lexical network, a few have investigated the underlying developmental process. In particular, Steyvers & Tenenbaum (2005) suggested that the observed effects of connectivity are the consequence of how the lexical network gets constructed in the child's mind. According to this explanation, known as Preferential Attachment, highly connected words in the child's lexicon tend to "attract" more words over time, in a rich-get-richer scenario (Barabasi & Albert, 1999). In other words, what predicts learning is the *internal* connectivity in the child's early lexicon. In contrast, Hills et al. (2009) suggested that what biases the learning is not the connectivity in the child's internal lexicon but, rather, *external* connectivity in the learning environment. They called this alternative explanation Preferential Acquisition. For clarity of reading, we will call preferential attachment the Internally-driven mechanism (INT), and preferential acquisition the Externally-driven mechanism (EXT). Figure 1 shows an illustration of both growth scenarios with the same simplified network.

These two proposals represent two divergent ideas about the role of lexical networks in acquisition. On the INT proposal, learning is driven by known words with high connectivity to other known words (Figure 1, left). Thus, the network structure is a causal factor in word learning, that is, children rely on the organization of their past knowledge to determine future learning (Altvater-Mackensen & Mani, 2013; Borovsky, Ellis, Evans, & Elman, 2016; Chi & Koeske, 1983; Storkel, 2009). In contrast, on the EXT approach, learning is driven by the connectivity of words that are not known yet (Figure 1, right). Thus, the relevant network structure is not internally represented by children, and the observed connectivity effect might be an epiphenomenon of some properties of the linguistic input. For example, highly connected concrete nouns in the input could be more easily learned because of their contextual diversity, allowing for easier meaning disambiguation (McMurray, Horst, & Samuelson, 2012; Smith & Yu, 2008; Yurovsky & Frank, 2015). Another reason could be

82 that these words are emphasized by the caregivers in their child-directed speech (Clark, 2007;

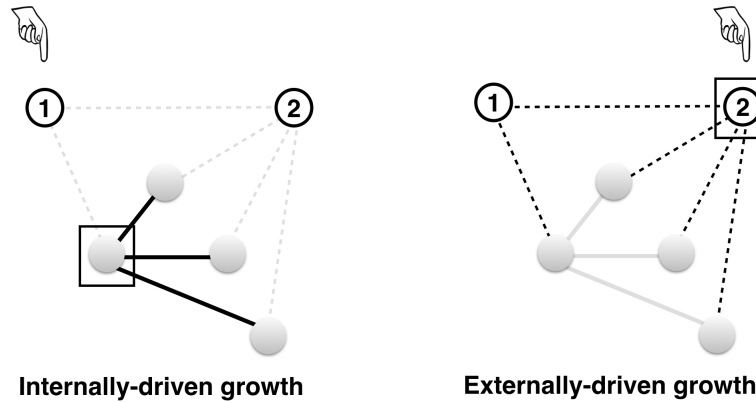83 Hoff & Naigles, 2002; Huttenlocher et al., 1991).

Figure 1. Illustration of the two growth scenarios. Filled grey circles represent known words (Internal) at a certain point in time. The empty, numbered circles represent words that have not yet been learned (External) and which are candidates to enter the lexicon next. The identity of the word that is going to be learned depends on the growth scenario. Here the squares indicate the node that drives growth in each scenario, and the hand pointer indicates which word is likely to be learned. For INT, the utility of a candidate external node is the average degree (i.e., number of links) of the internal nodes that it would attach to. In this simplified example, candidate node 1 would connect to an internal node with 3 connections; thus we have $u_{INT}(node_1) = 3$. As for candidate node 2, it would connect to internal nodes that have only one connection each, making an average of 1, i.e., $u_{INT}(node_2) = 1$. According to INT, node 1 is more likely to enter the lexicon. For EXT, the utility of a candidate node is its degree in the entire network. In our example, candidate node 1 has 2 connections in total, whereas candidate node 2 has 5 connections. So we have $u_{EXT}(node_1) = 2$ and $u_{EXT}(node_2) = 5$. Thus, according to EXT, node 2 is more likely to enter the lexicon next. This figure is based on an example from Hills et al. (2009).

84   Hills et al. (2009) investigated the growth of lexico-semantic networks in toddlers and

85 found that growth did not proceed according to INT as was originally hypothesized by

Steyvers & Tenenbaum (2005), but rather according to EXT.[1] This is an important finding

because it suggests that learning in the early stages is mostly driven by properties of the

external input, regardless of how past knowledge is organized. However, this work explored

the INT/EXT growth in a special case: networks that were based on 1) semantic

associations, 2) production-based vocabularies, and 3) data from English-learning children,

only. The extent to which this result depends on the domain (e.g., semantic vs. phonological

connectivity), the vocabulary measure (production vs. comprehension) and culture/language

is thus an open area for investigation (Hills & Siew, 2018). In this work, we test the

generality of prior findings along these three dimensions.

First, we study the phonological network in addition to the semantic network. These

two networks represent different ways the mental lexicon is structured (Beckage & Colunga,

2016). In particular, words that are neighbors in the semantic network (e.g., cat, dog) are

not necessarily neighbors in the phonological network and vice versa. Does the phonological

network also predict word learning? Previous work has found an effect of words' connectivity

in the phonological network on their age of learning (Carlson et al., 2014; Stella et al., 2017;

Storkel, 2009). In other words, words learned earlier in life tend to sound similar to many

other words than a word learned later in life. However, this finding is *a priori* compatible

with both INT and EXT, and previous studies did not explicitly compare these two

mechanisms. Here, we investigate whether phonological networks, like semantic networks,

grow through EXT, or if they rather grow via INT (Figure 1).

Second, we study vocabularies measured using both comprehension and production.

Previous studies have found differences between these vocabularies in terms of their content

---

[1] Besides INT and EXT, the authors tested a third mechanism (called the lure of associates) which resembles EXT in that it is driven by the connectivity of external nodes, except that this connectivity is computed with respect to words that are known. However, EXT is the externally-driven scenario that best predicted the data in this previous work.

and rate of acquisition (Bates, Dale, & Thal, 1995; Benedict, 1979; Fenson et al., 1994). These differences may reflect the fact that comprehension and production do not share the same constraints. For instance, whereas comprehension depends on the ease with which words are stored and accessed, production depends, additionally, on the ease with which words are articulated, e.g., shorter words are produced earlier (Braginsky et al., 2019). By investigating comprehension-based vocabularies, we assess the extent to which the network growth mechanism captures general learning patterns beyond the specific constraints of production.

Finally, we use developmental data in 10 languages. Lexical networks can show more or less cross-linguistic variability along both the semantic and phonological domains (Arbesman, Strogatz, & Vitevitch, 2010; Lupyan & Lewis, 2017; Youn et al., 2016). Besides, cultures might differ in the way caregivers talk to children (Cristia, Dupoux, Gurven, & Stieglitz, 2017; Kuhl et al., 1997), and this difference in the input could influence the way in which the children's networks grow. Thus, cross-linguistic comparison is crucial to test the extent to which growth mechanisms are equally engaged across a wider variety of cultures compared with the extent to which the growth mechanisms are specific to patterns of learning that emerge due to the particulars of a given language or culture (Bates & MacWhinney, 1987; Slobin, 2014).

We adopted the following research strategy. We used parent reports on the MacArthur-Bates Communicative Development Inventory and its cross-linguistic adaptations (Fenson et al., 1994; Frank, Braginsky, Yurovsky, & Marchman, 2017). We studied the timeline of word learning using the normative age of acquisition (i.e., the age at which at least 50% of children know a given word). Our choice of studying the normative learning trajectory rather than the individual trajectories was motivated by the nature of the dataset used—which is primarily based on cross-sectional studies. Children may vary in their individual learning trajectories, but the aggregate data provide highly robust measures of the

134 average learning patterns (Fenson et al., 1994). The use of such measures has lead to

135 important insights on the mechanisms of word learning (Goodman et al., 2008; Hills et al.,

136 2010, 2009; Stella et al., 2017; Storkel, 2009).

137 The paper is organized as follows. First, we describe the datasets we used and explain

138 how we constructed the networks. Second, we analyze static properties of words'

139 connectivity in these networks (correlation with age of acquisition and shape of the

140 distribution), and we explain how these properties inform hypotheses about network growth.

141 Next, we fit the two hypothesized growth mechanisms to the data. We investigate the extent

142 to which the results obtained in Hills et al. (2009) generalize to phonological networks and

143 comprehension-based vocabularies, and whether this generalization holds cross-linguistically.

Networks

Data

146 We used data from Wordbank (Frank et al., 2017), an open repository aggregating

147 cross-linguistic language developmental data of the MacArthur-Bates Communicative

148 Development Inventory (CDI), a parent report vocabulary checklist. Parent report is a

149 reliable and valid measure of children's vocabulary that allows for the cost-effective collection

150 of datasets large enough to test network-based models of acquisition (Fenson et al., 1994).

151 When filling out a CDI form, caregivers are either invited to indicate whether their child

152 "understands" (comprehension) or "understands and says" (production) each of about

153 400-700 words. For younger children (e.g., 8 to 18 months in the English data), both

154 comprehension and production are queried, whereas for older children (16 to 36 months) only

155 production is queried. Due to these limitations, we use data from younger children to test

156 comprehension and data from older children to test production. In addition, following

157 previous studies (Hills et al., 2009; Storkel, 2009), we restricted our analysis to the category

158 of nouns due to the fact that nouns predominate the early expressive and receptive lexicons

159 (Bates et al., 1995). Their larger sample size (compared, for example, to verbs or adjectives)

is more suited to the network-based analysis of development. Table 1 gives an overview of the data we used.

Age of acquisition

For each word in the CDI data, we compute the proportion of children who understand or produce the word at each month. Then we fit a logistic curve to these proportions and determined when the curve crosses 0.5, i.e., the age at which at least 50% of children know the word. We take this point in time to be each word's age of acquisition (Braginsky et al., 2019; Goodman et al., 2008).

Table 1

Statistics for the dataset we used. The ages are in months.

Language	Comprehension			Production		
	Nouns	Ages	N	Nouns	Ages	N
Croatian	209	8-16	250	312	16-30	377
Danish	200	8-20	2,398	316	16-36	3,714
English	209	8-18	2,435	312	16-30	5,520
French	197	8-16	537	307	16-30	827
Italian	209	7-24	648	312	18-36	752
Norwegian	193	8-20	2,922	316	16-36	9,303
Russian	207	8-18	768	314	18-36	1,037
Spanish	208	8-18	788	312	16-30	1,146
Swedish	205	8-16	467	339	16-28	900
Turkish	180	8-16	1,115	297	16-36	2,422

Semantic networks

We constructed semantic networks for English data following the procedure outlined in Hills et al. (2009), as follows. We used as an index of semantic relatedness the Florida Free Association Norms (Nelson et al., 1998). This dataset was collected by giving adult participants a word (the cue), and asking them to write the first word that comes to mind (the target). For example, when given the word "ball," they might answer with the word "game." A pair of nodes were connected by a directed link from the cue to the target if there was a cue-target relationship between these nodes in the association norms. The connectivity of a given node was characterized by its *indegree*: the number of links for which the word was the target.[2] To model growth from month to month, we constructed a different network at each month, based on the nouns that have been acquired by that month.

Since the free association norms are available only in English, we used the hand-checked translation equivalents available in Wordbank, which allowed us to use the English association norms across 10 languages. Semantic associations are not necessarily shared across languages, but we use this technique as a reasonable first approximation. In support of this approximation, Youn et al. (2016) showed that semantic networks across languages share substantial similarities. The fact that semantic associations are assumed to be shared across languages does not mean that the semantic networks will necessarily grow in a similar fashion. For instance, the set of words acquired by children as well as the order of word acquisition can vary from language to language leading to possibly different learning strategies.

---

[2] This choice was based on prior work by Hills et al. (2009) stating that analyses with both outdegrees (sum of the links where the word is the cue in a cue-target pair) and total degree (outdegree plus indegree) led to results weaker than those calculated with indegree.

Phonological networks

To construct phonological networks we first mapped the orthographic transcription of words to their International Phonetic Alphabet (IPA) transcriptions in each language, using the open-source text-to-speech software **Espeak**. This software provides the correct IPA transcription if the word is found in a spelling-to-phonemes dictionary, otherwise it uses language-specific pronunciation rules to generate an approximate phonetic transcription. We used the Levenshtein distance (also known as edit distance) as a measure of phonological relatedness between two nodes. The measure counts the minimum number of operations (insertions, deletions, substitutions) required to change one string into another.

In previous studies, two nodes were linked if they had an edit distance of 1 (Carlson et al., 2014; Stella et al., 2017; Storkel, 2009). However, these studies reported a contribution of phonological connectivity to noun learning when networks were built using a dense adult vocabulary. Since the focus of the current study is on the mechanism of growth, the networks are based on children's early vocabulary. The latter, however, contains very few noun pairs with an edit distance of 1. To better represent the similarity space in the phonological domain, we increased the threshold from 1 to 2, that is, two nodes were related if their edit distance was equal to 1 or 2.[3] The connectivity of a given node was characterized with its *degree*: the number of links it shares with other words.

---

[3] In Appendix A, we show the main results for phonological networks based on an edit distance of 1. We also show the results for phonological networks where the edges between pairs of words were weighted by a normalized edit distance. We did not consider the case of a threshold larger than 2 since many short pairs appear phonologically unrelated when the edit distance is 3 or more (e.g., "cat"/"dog").

₂₀₇   **Analysis**

Static properties of the global network

₂₀₉   We start by analyzing word connectivity in the global (static) network. We constructed

₂₁₀ this network using nouns learned by the oldest age for which we have CDI data (e.g., in

₂₁₁ English this corresponds, in comprehension, to the network by 18 months, and in production,

₂₁₂ to the network by 30 months). This global network is the end-state towards which both INT

₂₁₃ and EXT converge by the last month of learning. Moreover, following Hills et al. (2009), we

₂₁₄ used this end-state network as a proxy for the external connectivity in the learning

₂₁₅ environment. Below we analyze properties of these global networks that may a priori hint at

₂₁₆ an INT- or EXT-like growth.

₂₁₇   **Connectivity predicts the age of acquisition.**   Connectivity in the global

₂₁₈ network is directly related to EXT as it represents the explicit criterion this growth scenario

₂₁₉ uses to determine what words should be learned first (Figure 1). Therefore, a direct

₂₂₀ consequence of an EXT-like growth scenario is a correlation between connectivity in the

₂₂₁ global network and the age of acquisition. This correlation is also necessary to INT, although

₂₂₂ the causality is reversed: Higher connectivity in the global network is caused by earlier

₂₂₃ learning, not the other way around. Some words end up being highly connected in the global

₂₂₄ network precisely because they happen to be acquired earlier and, therefore, have a higher

₂₂₅ chance of accumulating more links over time. Thus, the correlation between connectivity in

₂₂₆ the end-state network and AoA can result from both EXT and INT. If there is no such

₂₂₇ correlation, neither growth scenario can be posited as a possible learning mechanism.

₂₂₈   Figures 2 and 3 show how the age of acquisition in production and comprehension,

₂₂₉ respectively, correlates with the degree (or indegree for the semantic networks). For ease of

₂₃₀ visual comparison, the predictor (i.e., the degree) was centered and scaled. The plots show,

₂₃₁ overall, a negative correlation between the month of acquisition and the degree. In

₂₃₂ production data, the average correlation across languages was -0.24 ($SD = 0.10$) for the
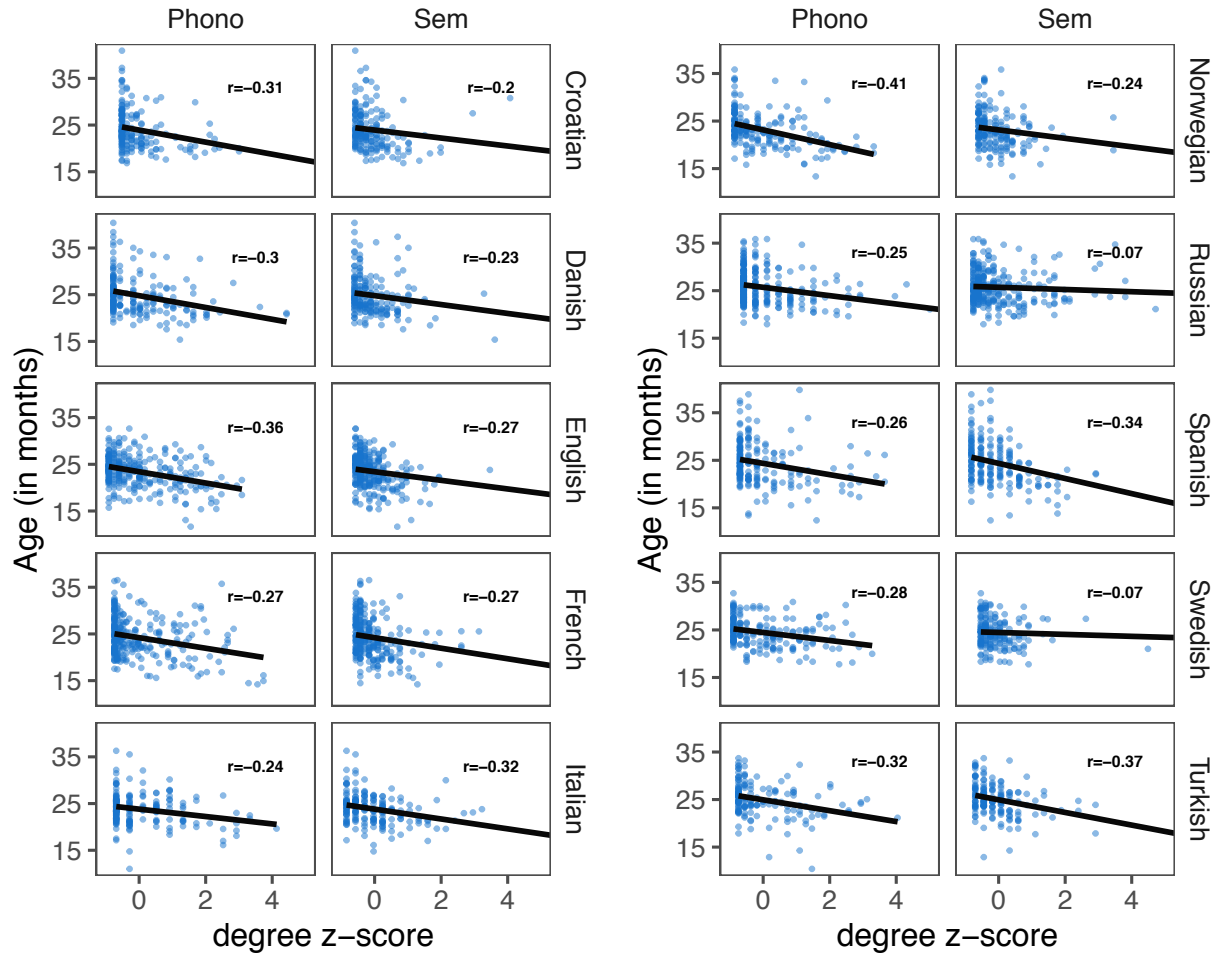
Figure 2. Production data (Age of acquisition) as predicted by the degree (i.e., connectivity) in this network. Results are shown in each language for phonological and semantic networks. Each point is a word, with lines indicating linear model fits, and numbers indicating the Pearson correlation coefficients.

semantic networks and -0.30 ($SD = 0.05$) for the phonological networks. In comprehension data, the average correlation was -0.21 ($SD = 0.08$) for the semantic networks and -0.21 ($SD = 0.07$) for the phonological networks. These results indicate that nouns with higher degrees are generally learned earlier, thus replicating previous findings in English (Hills et al., 2009; Storkel, 2009) and extending these findings to 10 different languages, generally, in both production- and comprehension-based vocabularies.
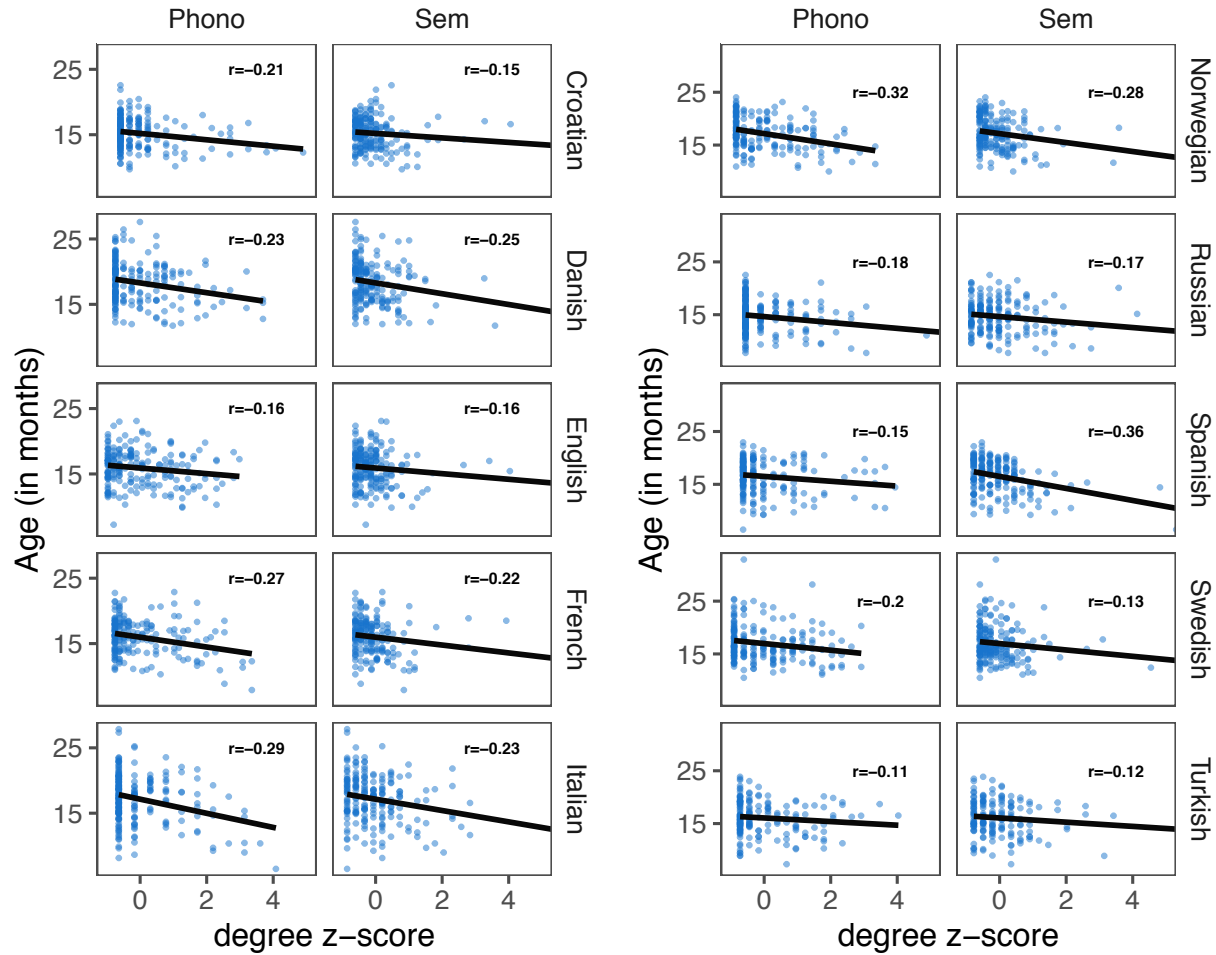
Figure 3. Comprehension data (Age of acquisition) as predicted by the degree (i.e., connectivity) in this network. Results are shown in each language for phonological and semantic networks. Each point is a word, with lines indicating linear model fits, and numbers indicating the Pearson correlation coefficients.

Power-law degree distribution. We also analyzed the global network's degree distribution. The shape of this distribution is particularly relevant to INT as this growth scenario is known to generate networks with a power-law degree distribution, i.e., a distribution of the form $p(k) \propto \frac{1}{k^\alpha}$ (Barabasi & Albert, 1999). If the end-state network displays this property, this fact would suggest, but not prove, an INT-like generative process. If, however, the degree distribution is very different from a power law, this would significantly weaken the case for INT. The log-log plots are shown in Figure 4. We fit a
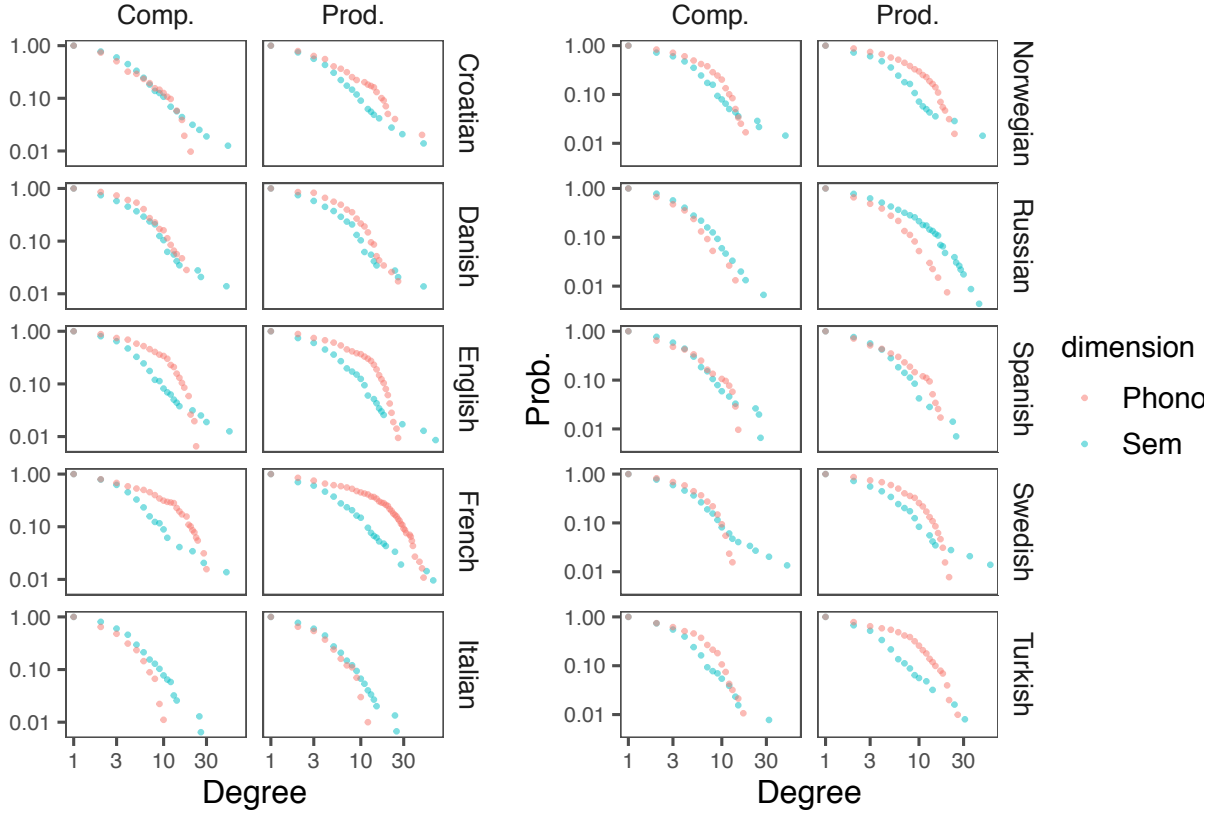
Figure 4. Log-log plot of the cumulative degree distribution function for the global phonological and semantic networks across languages. The figure shows the results for both production and comprehension data. A perfect power-law distribution should appear as a straight line in this graph.

power law to each empirical degree distribution following the procedure outlined in Clauset, Shalizi, & Newman (2009) and using a related R package (poweRlaw, Gillespie, 2015).

In brief, the analysis consisted in two steps. First, we derived the optimal cut-off, k_{min}, above which the distribution is more likely to follow a power-law,[4] and we estimate the corresponding scaling parameter α. Second, we calculated the goodness-to-fit, which resulted in a *p*-value quantifying the plausibility of the model. The results are shown in Table 2 for production data, and in Table 3 for comprehension data.

---

[4] In natural phenomena, it is often the case that the power law applies only for values above a certain minimum.

Table 2

Results of fitting a power-law model to the degree (i.e., connectivity) distribution in each model for production data. Numbers indicate the cut-off degree, the scaling parameter alpha, and the p-value which quantifies the plausibility of the power-law hypothesis. If the p-value is close to 1, a power law cannot be rejected as a plausible fit for the data.

	Sem.			Phono.		
Language	cut-off	alpha	p-value	cut-off	alpha	p-value
Croatian	4	2.55	0.881	4	2.18	0.123
Danish	4	2.38	0.001	11	4.55	0.858
English	5	2.66	0.132	20	9.14	0.511
French	8	2.81	0.133	20	3.75	0.112
Italian	4	2.93	0.608	9	9.45	0.780
Norwegian	5	2.88	0.201	15	6.28	0.744
Russian	24	5.61	0.723	8	4.20	0.541
Spanish	4	2.98	0.460	13	8.75	0.736
Swedish	4	2.49	0.171	11	4.68	0.103
Turkish	4	2.87	0.925	8	3.26	0.375

Overall, we could not reject the null hypothesis of a power-law distribution: The p-value was generally above 0.1 in almost all languages for both production and comprehension. That said, phonological networks had relatively larger cut-offs than semantic networks. These "truncated" power-laws in phonological networks may be due to the constraints that exist on word formation in the phonological domain such as the size of the phonemic inventory, phonotactic rules, and word length. Such constraints may limit the number of words that are phonologically similar, thus leading to distributions that decay faster than a non-truncated power law (Arbesman et al., 2010).

Table 3

Results of fitting a power-law model to the degree distribution in each model for comprehension data. Numbers indicate the cut-off degree, the scaling parameter alpha, and the p-value which quantifies the plausibility of the power-law hypothesis. If the p-value is close to 1, a power law cannot be rejected as a plausible fit for the data.

Language	Sem.			Phono.		
	cut-off	alpha	p-value	cut-off	alpha	p-value
Croatian	5	2.67	0.895	2	2.06	0.020
Danish	4	2.39	0.005	5	2.98	0.136
English	4	2.64	0.765	13	5.16	0.235
French	4	2.63	0.330	18	5.58	0.336
Italian	4	2.88	0.688	8	10.27	0.909
Norwegian	5	2.87	0.433	13	7.65	0.440
Russian	8	3.91	0.952	5	3.97	0.854
Spanish	5	3.11	0.552	5	3.01	0.085
Swedish	5	2.81	0.713	9	6.75	0.102
Turkish	4	3.13	0.887	9	5.73	0.958

In sum, the static properties of the end-state network are *a priori* compatible with both INT and EXT. In order to decide between these two developmental scenarios, we need to fit explicit growth models to the data.

Network growth models

To test the network growth scenarios, we fit two growth models to the data. We calculated the probability that a word w_i, with a utility value u_i would enter the lexicon at a
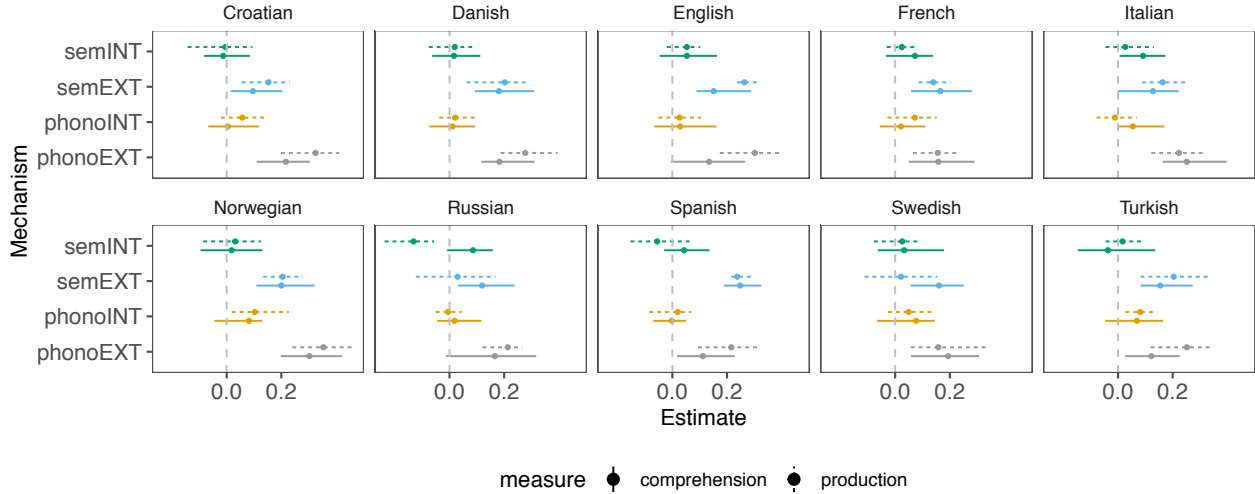
Figure 5. Evaluation of growth scenarios (EXT: externally-driven, INT: internally-driven) for both semantic and phonological networks. Each point represents the mean of the posterior distribution of the growth parameter, with ranges representing 95% credible intervals. Positive values mean that learning proceeds according to the predictions of the growth scenario, whereas negative values mean that learning proceeds in opposition to the predictions of the growth scenario.

given month, using a softmax function:

$$p(w_i) = \frac{e^{\beta u_i}}{\sum_j e^{\beta u_j}} \tag{1}$$

where β is a fitted parameter that captures the magnitude of the relationship between network parameters and growth (analogous to a regression coefficient). A positive value of β means that words with higher utility values u_i are acquired first, and a negative value means that words with lower utility values are acquired first (see Figure 1 for an illustration of how utility values u_i are defined in each growth scenario). The normalization includes all words that could be learned at that month.

We estimated the parameter β using a Bayesian approach. The inference was performed using the probabilistic programming language WebPPL (Goodman & Stuhlmuller, 2014). We defined a uniform prior over β, and at each month, we computed the likelihood

277 function over words that could possibly enter the lexicon at that month, fit to the words that

278 have been learned at that month (using Formula 1). Markov Chain Monte Carlo sampling

279 resulted in a posterior distribution over β, which we summarized in Figure 5. The results

280 replicate Hills et al.'s original finding regarding the semantic network in English and the

281 production-based vocabulary, which is that this network grows by EXT, not by INT.

282 Crucially, our results show that, generally speaking, this finding generalizes to

283 comprehension-based vocabulary, and holds across languages. This generalization was

284 obtained in both the semantic[5] and phonological domains. In Appendix B, we show that the

285 semantic and phonological domains provide largely non-redundant information.

Comparison to other predictors of age of acquisition

287   Above we showed that the way semantic and phonological information is structured in

288 the learning environment contributes to noun learning (via EXT) across languages. However,

289 we know that other factors influence learning as well (e.g., Braginsky et al., 2019). Next, we

290 investigated how semantic and phonological connectivity interact with two other factors.

291 The first one is word frequency, a well-studied factor shown to predict the age of acquisition

292 in a reliable fashion (e.g., Goodman et al., 2008). The second factor is word length, which

293 was shown to correlate with phonological connectivity: Shorter words are more likely to have

294 higher connectivity (Pisoni, Nusbaum, Luce, & Slowiaczek, 1985; Vitevitch & Rodrı'guez,

295 2005).

296   Since we found INT to be uninformative, we dropped it from this analysis, keeping

---

[5] One could imagine that the fact of using English free association norms cross-linguistically would decrease the effect of non-English semantic networks because of possible cultural differences. However, our findings do not support this assumption; rather, it supports our initial approximation about the shared nature of the semantic similarity measure. That said, this approximation is not perfect. For example, there is evidence that a small part of the variance in free association data can be explained by phonological similarity (Kachergis, Cox, & Jones, 2011; Matusevych & Stevenson, 2018), thus leading to possibly minor cross-linguistic differences.
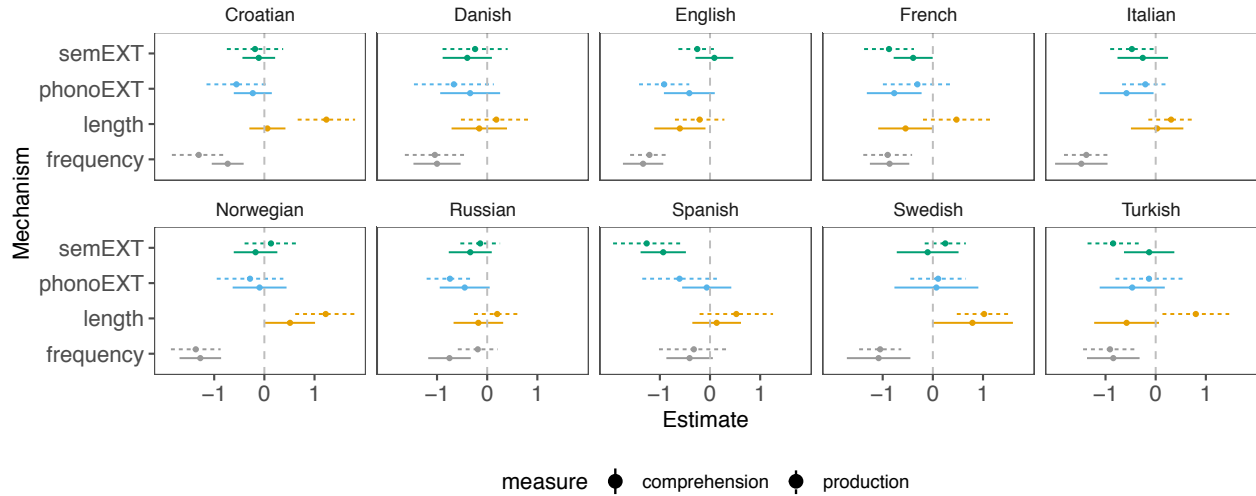
Figure 6. Estimates of the relative contribution of each predictor of AoA in the regression model in each language. Results are shown for both production and comprehension data. Ranges indicate 95% confidence intervals. Positive values indicate a positive relationship (e.g. longer words tend to have a higher AoA), while negative values indicate a negative relationship (e.g. words with higher frequency tend to have a lower AoA).
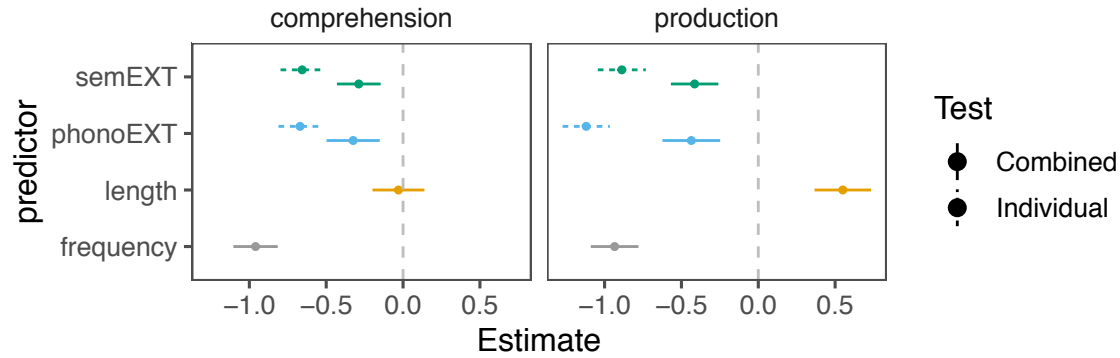

Figure 7. Estimates of the relative contribution of each predictor of AoA in the combined mixed-effects model with language as a random effect. Results are shown for both production and comprehension data. Ranges indicate 95% confidence intervals. Dotted ranges indicate the estimates for the predictor in a separate model that includes only this predictor as a fixed effect.

297 only EXT. This simplified the model because we no longer needed to fit growth

298 month-by-month. The latter was a requirement only for INT where the words' utilities

varied from month to month, depending on how connectivity changed in the growing internal
network. A more direct way to assess and compare the contribution of EXT in relation to
other word-level factors is through conducting regressions, where connectivity in the learning
environment, frequency, and length predict the age of acquisition.

For word length, we counted the number of phonemes in our generated IPA
transcription. For word frequency, we used the frequency estimates from Braginsky et al.
(2019) where unigram counts were derived based on CHILDES corpora in each language
(MacWhinney, 2014). Although these frequency counts use transcripts from independent sets
of children, they are based on large samples, and this allows us to average out possible
differences between children and the specificities of their input (see Goodman et al., 2008 for
a similar research strategy).

We conducted two analyses. We fit a linear regression for each language, and we fit a
linear mixed-effect model to all the data pooled across languages, with language as a random
effect. Figure 6 shows the coefficient estimate for each predictor in each language for
production and comprehension data. Figure 7 shows the coefficient estimates for all
languages combined (all predictors were centered and scaled).

The findings for the new predictors were as follows. Overall, frequency is the largest
and most consistent predictor of age of acquisition in both comprehension and production
data and across languages, endorsing results for nouns across a variety of analyses (Braginsky
et al., 2019; Goodman et al., 2008; Roy et al., 2015). Word length is more predictive for
production than comprehension (and this difference is very clear in the global model),
replicating previous work (Braginsky et al., 2019). Thus, word length seems to reflect the
effects of production's constraints rather than comprehension's constraints, i.e., longer words
are harder to articulate but they may not be significantly more difficult to store and access.

As for the factors of interest, i.e., semantic and phonological connectivity, we found

cross-linguistic differences. Connectivity contributes to learning in some languages but not in others. In particular, semantic connectivity does not explain variance in English data beyond that explained by phonological connectivity, frequency, and length. This finding contrasts with the original finding in Hills et al. (2009). However, this difference might be due to our using a slightly different model (which included word length as a covariate) and a larger dataset. That said, and despite these apparent cross-linguistic differences, both phonological and semantic connectivity are significant predictors in the combined model (Figure 7).

Discussion

This study provided an analysis of network growth during development. We compared two network growth scenarios described in the pioneering work of Steyvers & Tenenbaum (2005) and Hills et al. (2009). The first scenario, INT (originally called Preferential Attachment), described a rich-get-richer network growth model in which the current structure of the learner's internal network determines future growth; the other, EXT (originally called Preferential Acquisition) described a model in which the external, global environmental network structure determines learners' growth patterns. These two mechanisms represent two fundamentally different accounts of lexical growth: One suggests that future word knowledge is primarily shaped by the children's past knowledge and its organization, whereas the other suggests that learning is shaped, rather, by salient properties in the input regardless of how past knowledge is organized. The present study tested the generality of previous findings by 1) investigating phonological networks together with semantic networks, 2) testing both comprehension- and production-based vocabularies, and 3) comparing the results across 10 languages.

We found that the original findings reported in Hills et al. (2009) generalize well across all these dimensions. First, just like semantic networks, phonological networks grow via the externally-driven scenario (EXT), not by the internally-driven mechanism (INT). Second, comprehension-based vocabularies grow in a way similar to production-based vocabularies.

Finally, the findings were, overall, similar across the 10 languages we tested. Although we find some cross-linguistic variation when semantic and phonological networks were pitted against frequency and length, this variability is to be taken with a grain of salt as it might be exaggerated in our study by several factors such as the limited and partially-overlapping set of nouns for each language, measurement error due to the sample of acquisition data, the sample of frequency data, and the translation of association norms. In fact, both phonological and semantic connectivity are significant predictors above and beyond frequency and length when data are pooled across languages.

These findings corroborate the hypothesis that children start by learning words that have high similarity to a variety of other words in the learning environment, not in the child's available lexicon. This hypothesis implies that children are sensitive to highly connected words although they do not initially have access to the full network, thus raising some important questions: What mechanism allows children to distinguish highly connected words from other words? Besides, why would highly connected words be easier to learn?

One possibility is that these patterns emerge from children's use of statistical learning abilities (Aslin & Newport, 2012; Saffran, Aslin, & Newport, 1996; Smith & Yu, 2008). The term "statistical learning" has been used in the developmental literature to describes the process by which one acquires information about their environment through tracking the frequency distribution of some elements (e.g., words) in different contexts. An important property of this kind of learning is that it occurs without explicit instructions and through mere exposure to the input. Previous work in this line of research has documented specific mechanisms that can explain the patterns found in the current study.

For example, in the semantic domain, growth according to EXT could be explained by a mechanism similar to cross-situational learning (McMurray et al., 2012; Smith & Yu, 2008; Yurovsky & Frank, 2015). According to this mechanism, children track the co-occurrence of concrete nouns with their possible semantic referents. The referent of a word heard in only

376 one naming situation can be ambiguous (e.g., when the word "ball" is heard for the first time

377 in the presence of both a ball and a chair), but hearing the same word in a diversity of

378 semantic contexts allows the learner to narrow down the set of possible word-object

379 mappings. In our case, free association (used to determine semantic network connectivity) is

380 related to contextual co-occurrence (Fourtassi & Dupoux, 2013; Griffiths, Steyvers, &

381 Tenenbaum, 2007), meaning that highly connected words will tend to occur in a variety of

382 speech and referential contexts. This fact makes such words easier to learn because they

383 have more referential disambiguating cues across learning contexts. Crucially, children can

384 learn these words without necessarily knowing the meaning of all other words with which

385 they co-occur (Fourtassi, Dunbar, & Dupoux, 2014), hence the similarity with EXT. This

386 possibility is supported by the finding that words' diversity of occurrence in child-directed

387 speech predicts their age of learning (Hills et al., 2010; Stella et al., 2017).

388   In the phonological case, network growth according to EXT is also compatible with a

389 scenario whereby children are tracking low-level statistical patterns, e.g., high probability

390 sound sequences. Indeed, connectivity in the phonological network is inherently correlated

391 with phonotactic probability (Vitevitch, Luce, Pisoni, & Auer, 1999). That is, highly

392 connected words tend to be made of frequent sound sequences. Children are sensitive to

393 local phonotactic regularities (Jusczyk, Luce, & Charles-Luce, 1994), and this sensitivity

394 might lead them to learn higher-probability words more easily (Storkel, 2001). This

395 explanation is supported by computational simulations that show how learning general

396 phonotactics patterns create "well-worn paths" which allow the models to represent several

397 distinct but phonologically neighboring words (Dell, Juliano, & Govindjee, 1993; Siew, 2013;

398 Takac, Knott, & Stokes, 2017). More generally, there is a growing interest in investigating

399 precisely how the local patterns acquired through statistical learning may give rise to the

400 global network organization (For a review, see Karuza, Thompson-Schill, & Bassett, 2016).

401   Besides using their own statistical learning skills, children could also benefit from the

way their caregivers speak. Perhaps the caregivers put more emphasis on the words that are highly connected in *their* mature lexical network. This emphasis would guide children to learn first these highly connected words, even though children do not have access to the distribution of words' connectivity in the final network. Investigating this possibility would require further research on caregiver-child interaction (MacWhinney, 2014; Roy et al., 2015), examining what words are introduced over development and the extent to which children's uptake is influenced by this input (Clark, 2007; Hoff & Naigles, 2002; Huttenlocher et al., 1991).

This study investigated the class of nouns in isolation — following previous studies investigating the early semantic and phonological network (Hills et al., 2009; Storkel, 2009). We could ask if studying one class separately is a legitimate research strategy. In other words, would word classes (such as nouns, verbs and function words) be acquired relatively differently, or would they interact substantially to the extent that it becomes unreasonable to study each class separately?

There are many observations that support the hypothesis that different word classes have different pathways of learning, making it worthwhile to study each class separately. For instance, different word classes follow different time trajectories: In the early stages of development, nouns tend to be acquired at a higher pace than predicates and function words (Bates et al., 1994). Research has shown that this difference cannot be trivially attributed to differences in the degree to which these classes are present in the input; if anything, verbs and function words are often more frequent in the input than nouns (e.g., Gentner, 1982). Goodman et al. (2008) found an effect of frequency on the age of acquisition within — not across — classes. Further, recent work by Braginsky et al. (2019) tested a large number of predictors, besides frequency, and found that these predictors do not influence the acquisition of word classes in the same way. For example, the acquisition of nouns was found to be most influenced by frequency and concreteness, whereas the acquisition of function

428   words was most influenced by word length.

429   This work shares a number of limitations with previous studies using similar research
430   strategy and datasets. Chief among these limitations is the fact that the age of word
431   acquisition is computed using different children at different ages (due to the fact that
432   available CDI data is mainly cross-sectional). Such a measure has been shown to be valid
433   and reliable (Fenson et al., 1994), and has allowed researchers to study important aspects of
434   word learning (Braginsky et al., 2019; Goodman et al., 2008; Hills et al., 2009; Stella et al.,
435   2017; Storkel, 2009). In our case, the use of cost-effective cross-sectional data has allowed us
436   to leverage large-scale studies across several languages. That said, it is important to
437   remember that this type of data can only inform us about the learning trajectory of the
438   "average" child. Although our study endorses, overall, the externally-driven account of
439   network growth, this does not mean individual children never use some variant of INT or
440   some combination of both INT and EXT (Beckage & Colunga, 2019). To illustrate, some
441   children develop "islands of expertise," that is, well-organized knowledge about a certain
442   topic (e.g., birds or dinosaurs). This prior knowledge enables these children to learn new
443   related words more easily (e.g., Chi & Koeske, 1983).

444   To conclude, our work validates and generalizes previous results in early network
445   development. It suggests that the advantage of highly connected words may result, at least
446   in the early stages of word learning, from the operation of simpler mechanisms in both the
447   semantic and phonological domains. One question for future experimental work is whether
448   such correlational patterns of growth can be produced in controlled behavioral experiments.

All data and code for these analyses are available at
https://github.com/afourtassi/networks

449

Acknowledgements

Disclosure statement

None of the authors have any financial interest or a conflict of interest regarding this work and this submission.

References

Altvater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech segmentation. *Developmental Science*, *16*(6).

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*(03), 679–685.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3).

Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language*. Oxford, England: Blackwell.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Erlbaum.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., . . . Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*(1).

Beckage, N. M., & Colunga, E. (2016). Language networks as models of cognition: Understanding cognition through language. In *Towards a theoretical framework for analyzing complex linguistic networks* (pp. 3–28). Springer.

Beckage, N. M., & Colunga, E. (2019). Network growth modeling to capture individual lexical learning. *Complexity*.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language*, *6*(2), 183–200.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, *19*(6).

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children?S word learning across languages. *Open Mind*, *3*.

Carlson, M. T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language*, *75*, 159–180.

Chi, M. T., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, *19*(1).

Clark, E. V. (2007). Young children's uptake of new words in conversation. *Language in Society*, *36*(2).

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6).

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive*

548   *Science, 17*(2), 149–195.

549   Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J.

550   (1994). Variability in early communicative development. *Monographs of the Society*

551   *for Research in Child Development, 59*(5).

552   Fourtassi, A., Dunbar, E., & Dupoux, E. (2014). Self-consistency as an inductive bias in

553   early language acquisition. In *Proceedings of the annual meeting of the cognitive*

554   *science society.*

555   Fourtassi, A., & Dupoux, E. (2013). A corpus-based evaluation method for distributional

556   semantic models. In *51st annual meeting of the association for computational*

557   *linguistics proceedings of the student research workshop* (pp. 165–171).

558   Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open

559   repository for developmental vocabulary data. *Journal of Child Language, 44*(3),

560   677–694.

561   Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural

562   partitioning. *Center for the Study of Reading Technical Report.*

563   Gillespie, C. S. (2015). Fitting heavy tailed distributions: The poweRlaw package. *Journal*

564   *of Statistical Software, 64*(2), 1–16. Retrieved from http://www.jstatsoft.org/v64/i02/

565   Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the

566   acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531.

567   Goodman, N., & Stuhlmuller, A. (2014). The Design and Implementation of Probabilistic

568   Programming Languages. http://dippl.org.

569   Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation.

570   *Psychological Review, 114*(2), 2007.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739.

Hills, T. T., & Siew, C. S. (2018). Filling gaps in early word learning. *Nature Human Behaviour*, *2*(9).

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2).

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2).

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infant's sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645.

Kachergis, G., Cox, G. E., & Jones, M. N. (2011). OrBEAGLE: Integrating orthography into a holographic model of the lexicon. In *International conference on artificial neural networks* (pp. 307–314). Springer.

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: Influences of network topology on human learning. *Trends in Cognitive Sciences*, *20*(8).

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., . . . Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

594 Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation

595   model. *Ear and Hearing, 19*(1).

596 Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues: The role of

597   language in semantic knowledge. *Language, Cognition and Neuroscience.*

598 MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II.*

599   Psychology Press.

600 Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science,*

601   *14*(1), 57–77.

602 Matusevych, Y., & Stevenson, S. (2018). Analyzing and modeling free word associations. In

603   *Proceedings of the 40th Annual Conference of the Cognitive Science Society.*

604 McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the

605   interaction of online referent selection and slow associative learning. *Psychological*

606   *Review, 119.*

607 McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature

608   production norms for a large set of living and nonliving things. *Behavior Research*

609   *Methods, 37*(4), 547–559.

610 Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida

611   word association, rhyme, and word fragment norms. Retrieved from

612   http://w3.usf.edu/FreeAssociation/

613 Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception,

614   word recognition and the structure of the lexicon. *Speech Communication, 4*(1),

615   75–95.

616 Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of

617   a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41).

618 Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old

619   infants. *Science*, *274*(5294), 1926–1928.

620 Siew, C. S. (2013). Community structure in the phonological network. *Frontiers in*

621   *Psychology*, *4*.

622 Slobin, D. I. (2014). *The crosslinguistic study of language acquisition* (Vol. 4). Psychology

623   Press.

624 Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via

625   cross-situational statistics. *Cognition*, *106*(3).

626 Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns

627   in early word acquisition in children. *Scientific Reports*, *7*.

628 Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks:

629   Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78.

630 Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development.

631   *Journal of Speech, Language, and Hearing Research*, *44*(6), 1321–1337.

632 Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and

633   semantic variables on word learning by infants. *Journal of Child Language*, *36*(2),

634   29–321.

635 Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning

636   of specific english words. *Child Development*, *89*(4).

637 Takac, M., Knott, A., & Stokes, S. (2017). What can neighbourhood density effects tell us

638   about word learning? Insights from a connectionist model of vocabulary development.

639   *Journal of Child Language*, *44*(2).

640 Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical

641   retrieval? *Journal of Speech, Language, and Hearing Research*, *51*(2), 408–422.

642 Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics,

643   neighborhood activation, and lexical access for spoken words. *Brain and Language*,

644   *68*(1), 306–311.

645 Vitevitch, M. S., & Rodrı'guez, E. (2005). Neighborhood density effects in spoken word

646   recognition in spanish. *Journal of Multilingual Communication Disorders*, *3*(1).

647 Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., . . . Bhattacharya,

648   T. (2016). On the universal structure of human lexical semantics. *Proceedings of the*

649   *National Academy of Sciences*, *113*(7), 1766–1771.

650 Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on

651   cross-situational learning. *Cognition*, *145*.

652 Altvater-Mackensen, N., & Mani, N. (2013). Word-form familiarity bootstraps infant speech

653   segmentation. *Developmental Science*, *16*(6).

654 Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological

655   networks across multiple languages. *International Journal of Bifurcation and Chaos*,

656   *20*(03), 679–685.

657 Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to

658   forming general rules. *Current Directions in Psychological Science*, *21*(3).

659 Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*,

660   *286*(5439), 509–512.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language*. Oxford, England: Blackwell.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Erlbaum.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., . . . Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language, 21*(1).

Beckage, N. M., & Colunga, E. (2016). Language networks as models of cognition: Understanding cognition through language. In *Towards a theoretical framework for analyzing complex linguistic networks* (pp. 3–28). Springer.

Beckage, N. M., & Colunga, E. (2019). Network growth modeling to capture individual lexical learning. *Complexity*.

Benedict, H. (1979). Early lexical development: Comprehension and production. *Journal of Child Language, 6*(2), 183–200.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science, 19*(6).

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children?S word learning across languages. *Open Mind, 3*.

Carlson, M. T., Sonderegger, M., & Bane, M. (2014). How children explore the phonological network in child-directed speech: A survival analysis of children's first word productions. *Journal of Memory and Language, 75*, 159–180.

Chi, M. T., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19*(1).

Clark, E. V. (2007). Young children's uptake of new words in conversation. *Language in Society, 36*(2).

Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review, 51*(4), 661–703.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6).

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development.*

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17*(2), 149–195.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5).

Fourtassi, A., Dunbar, E., & Dupoux, E. (2014). Self-consistency as an inductive bias in early language acquisition. In *Proceedings of the annual meeting of the cognitive science society.*

Fourtassi, A., & Dupoux, E. (2013). A corpus-based evaluation method for distributional semantic models. In *51st annual meeting of the association for computational linguistics proceedings of the student research workshop* (pp. 165–171).

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language, 44*(3), 677–694.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *Center for the Study of Reading Technical Report.*

Gillespie, C. S. (2015). Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software, 64*(2), 1–16. Retrieved from http://www.jstatsoft.org/v64/i02/

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language, 35*(3), 515–531.

Goodman, N., & Stuhlmuller, A. (2014). The Design and Implementation of Probabilistic Programming Languages. http://dippl.org.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114*(2), 2007.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language, 63*(3), 259–273.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science, 20*(6), 729–739.

Hills, T. T., & Siew, C. S. (2018). Filling gaps in early word learning. *Nature Human Behaviour, 2*(9).

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2).

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2).

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infant's sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*(5), 630–645.

Kachergis, G., Cox, G. E., & Jones, M. N. (2011). OrBEAGLE: Integrating orthography into a holographic model of the lexicon. In *International conference on artificial neural networks* (pp. 307–314). Springer.

Karuza, E. A., Thompson-Schill, S. L., & Bassett, D. S. (2016). Local patterns to global architectures: Influences of network topology on human learning. *Trends in Cognitive Sciences*, *20*(8).

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, *277*(5326), 684–686.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, *19*(1).

Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues: The role of language in semantic knowledge. *Language, Cognition and Neuroscience*.

MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk, Volume II*. Psychology Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

Matusevych, Y., & Stevenson, S. (2018). Analyzing and modeling free word associations. In *Proceedings of the 40th Annual Conference of the Cognitive Science Society*.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119.*

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*(4), 547–559.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from http://w3.usf.edu/FreeAssociation/

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*(1), 75–95.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences, 112*(41).

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928.

Siew, C. S. (2013). Community structure in the phonological network. *Frontiers in Psychology, 4.*

Slobin, D. I. (2014). *The crosslinguistic study of language acquisition* (Vol. 4). Psychology Press.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3).

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns

in early word acquisition in children. *Scientific Reports, 7*.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*(1), 41–78.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*(6), 1321–1337.

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language, 36*(2), 29–321.

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific english words. *Child Development, 89*(4).

Takac, M., Knott, A., & Stokes, S. (2017). What can neighbourhood density effects tell us about word learning? Insights from a connectionist model of vocabulary development. *Journal of Child Language, 44*(2).

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research, 51*(2), 408–422.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*(1), 306–311.

Vitevitch, M. S., & Rodrı'guez, E. (2005). Neighborhood density effects in spoken word recognition in spanish. *Journal of Multilingual Communication Disorders, 3*(1).

Youn, H., Sutton, L., Smith, E., Moore, C., Wilkins, J. F., Maddieson, I., . . . Bhattacharya, T. (2016). On the universal structure of human lexical semantics. *Proceedings of the National Academy of Sciences, 113*(7), 1766–1771.

799   Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on

800   cross-situational learning. *Cognition, 145*.