Word Learning as Network Growth: A Cross-linguistic Analysis

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

[1] Department of Psychology, Stanford University

[2] Department of Psychology, University of Illinois

Author Note

Abdellah Fourtassi

Department of Psychology

Stanford University

50 Serra Mall

Jordan Hall, Building 420

Stanford, CA 94301

Correspondence concerning this article should be addressed to Abdellah Fourtassi, 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 both the phonological and semantic dimensions. Though this connectivity effect has been extensively documented, little is known about the underlying developmental mechanism. One view suggests that learning is primarily driven by a network growth model where highly connected words in the child's early lexicon attract similar words. Another view suggests 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, and across 8 languages. We show that external connectivity in the learning environment drives growth in both the semantic and the phonological networks, and that this pattern is consistent cross-linguistically. The findings suggest a word learning mechanism where children harness their statistical learning abilities to (indirectly) detect and learn highly connected words in the learning environment.

Keywords: Language understanding; audio-visual processing; word learning; speech perception; computational modeling.

Word Learning as Network Growth: A Cross-linguistic Analysis

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., B. C. Roy, Frank, DeCamp, Miller, & Roy, 2015). One strategy consists in documenting the timeline of words' acquisition, and studying the properties that make words easy or hard to learn. For example, within a lexical category, words that are more frequent in child-directed speech are acquired earlier (J. C. Goodman, Dale, & Li, 2008). Other factors include word length, the mean length of utterances in which the word occurs, and concreteness (see Braginsky, Yurovsky, Marchman, & Frank, 2016).

Besides these word-level properties, the lexical structure (that is, how words relate to each other) also influences the age of acquisition 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 and Loftus, 1975). Previous studies have investigated early vocabulary structure by constructing networks using a variety of word-word relations including shared semantic features, target-cue relationships in free association norms, co-occurrence in child directed speech, and phonological similarity. These studies have 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 (Engelthaler & Hills, 2017; Hills, Maouene, Riordan, & Smith, 2010; Hills, Maouene, Maouene, Sheya, & Smith, 2009; Stella, Beckage, & Brede, 2017; Storkel, 2009) add Beckage, 2011 and other people…

While most studies have focused on the static properties of the lexical network, a few have investigated the underlying developmental process. In particular, Steyvers and 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 word 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, network structure is a causal factor in early word learning; in contrast, on the EXT approach, network structure is not internally represented and, therefore, might be an epiphenomenon of the statistics of the linguistic input.

Hills et al. (2009) found that early lexical networks do not grow through INT as was originally hypothesised by Steyvers and Tenenbaum (2005), but rather through EXT. This finding, though potentially very profound, has only been estabished in the special case of networks that are based on 1) semantic associations, 2) word production as a measure of acquisition, and 2) data from English-learning children, only. The extent to which this result depends on the domain, the measure and culture/languagae is still unclear. In this work, we test the generality of the finding along these three dimesnions.

First, we study the phonological network in addition to the semantic network. These two networks represent different ways the mental lexicon is strcutured. 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 influence word learning? Previous work did find an effect of words' connectivity in the phonological network on their age of learning (Storkel, 2004, 2009; Stokes 2014). In other words, words

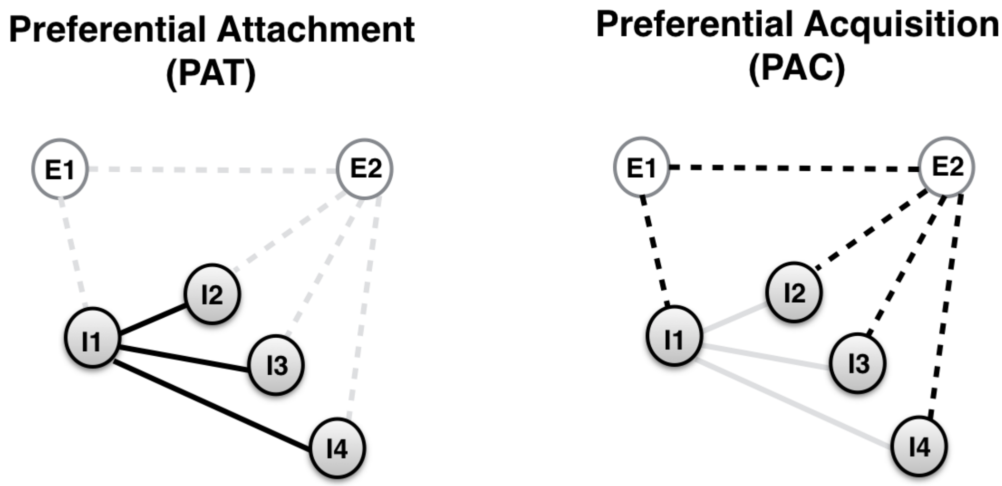

Figure 1. Illustration of the two growth scenarios. Filled circles (I1-I4) represent known words (Internal), and empty circles (E1 and E2) represent words that have not been learned yet (External). Black lines represent links that are relevant in each growth scenario, and gray lines represent links that are irrelevant. 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. Thus, according to INT, the node E1 is more likely to enter the lexicon first. For EXT, the utility of a candidate node is its degree in the entire network. According to EXT, the node E2 is more likely to enter the lexicon first.

learned earlier in life tend to sound similar to many other words than a word learned later in life. However, as mentioned above, this finding did not examine the underlying learning mechanism.[1] 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 and rate of acquistion (Benedict, 1979; Fensen et al. 1993). These differences can be due to the fact that comprehension and production do not share the same contraints. For instance, whereas comprehension depends on the ease with which words are stored and accessed,

[1]In particular, this finding is compatible with both INT and EXT.

production depends additionally on the ease with witch words are articuated (e.g., shorter words are produced earlier, Braginsky et al. 2016). By including comprehension-based vocabularies in our analysis, we assess the extent to which the network growth mechanisms capture 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-lingustic variability along both the semantic and phonological domains (Youn et al. 2015; Arbesman et al. 2009).[2] Besides, cultures might differ in the way caregivers guide the chidlren's learning patterns or in the way the adults' linguistic input is organized (Kuhl et al., 1997; Cristia et al. 2017), and this difference could influecne the learning mechanism. Thus, Cross-linguistic comparison is crucial to test what mechanism is cognitively universal and is used by all children, and what mechansism is specific to some patterns of learning that emerge due to the particulars of a given language or culture (Bates & MacWhinney, 1987; Slobin, 1985).

HERE ORGANIZATION OF THE PAPER

Networks

Data

We used data from Wordbank (Frank, Braginsky, Yurovsky, & Marchman, 2017), an open repository aggregating cross-linguistic language developmental data of the MacArthur-Bates Communicative Development Inventory (CDI), a parent report vocabulary checklist. Parent report is a reliable and valid measure of children's vocabulary that allows for the cost-effective collection of datasets large enough to test network-based models of acquisition (Fenson et al., 1994). When filling out a CDI form, caregivers are either invited to indicate whether their child "understands" (comprehension) or "understands and says" (production) for each of about 400 words. For youner chidlren (between 8 and 18 months),

[2] The difference is more obvious in the phonolgical case where the network strcuture—and the distribution of word connectivity—can a priori change from language to language depending on various languisitc conventions.

both comprehension and production are queried, whereas for oder chidlren (between 16 to 36 months) only production is queried. We use data from younger chidlren to test comprehension and data from older chidlren to test production. Following previous studies (Hills et al., 2009; Storkel, 2009), we restrict our analysis to nouns. Table 1 gives an overview of the data source.

Age of acquisition

For each word in 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 lest 50% of chidlren know the word. We take this point in time to be each word's age of acquisition (J. C. Goodman et al., 2008) and Braginsky.

	language	total	translated	normed
1	Croatian	253	177	170
2	Danish	295	198	187
3	English	296	296	274
4	Italian	311	203	194
5	Norwegian	305	193	186
6	Russian	311	311	285
7	Spanish	240	173	163
8	Turkish	293	175	164

Table 1

Total number of nouns produced by toddlers in the CDI (left). We included in our study the subset of these nouns that had available English translations (middle). The final set consisted of nouns that had both available translations as well entries in the Free Association Norms (right).

Semantic networks

We constructed semantic networks for English data following the procedure outlined in Hills et al. (2009). We used as an index of semantic relatedness the Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 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. To model growth from month to month, we constructed a different network at each month, based on the words that have been acquired by that month.

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 languages. Note that this does not necessarily mean that the resulting networks will be the same across languages, or that these networks will grow similarly. Though using the same association data acoss langauges assumes that the semantic similarity measure is universal (which is a reasonable approximation, Youn et al. 2015), the set of words acquired by children as well as the timeline of this acquisition can still vary from language to language leading to possbility different learing strategies.

Phonological networks

To construct phonological networks we first mapped the orthogrphic transcrition of words to their International Phonetic Alphabet (IPA) transcriptions, across languages, using the open source text-to-speech software **Espeak.** 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 (e.g., Storkel, 2009). However, in these previous studies the network was built using an adult vocabulary. In the current study, however, network growth models are based on the children's early vocabulary which contains very few word pairs with an edit distance of 1. When using this threshold, the resulting networks were too sparse and uninformative. Thus, we increased the threshold from 1 to 2, that is, two nodes were related if their edit distance was equal to 1 or 2. The connectivity of a given node was characterized with its *degree*: the number of links it shares with other words.

Analysis

Static properties of the global network

We start by analyzing word connectivity in the global (static) network. We constructed this network using nouns learned by the oldest age for which we have CDI data (e.g., in English this corresponds, in comprehension, to the network by 18 months, and in production, to the network by 30 months). This global network is the end-state towards which both INT and EXT converge by the last month of learning. Moreover, following Hills et al. (2009), we used this end-state network as a proxy for the external connectivity in the learning environment. Below we analyze properties of this global networks that are relevant to INT and/or EXT.

Connectivity predicts the age of acquisition. Connectivity in the global network is directly related to EXT as it represents the explicit criterion this growth scenario uses to determine what words should be learned first (Figure 1). Therefore, a direct consequence of a EXT-like growth scenario is a correlation between connectivity in the global network and the age of acquisition.[3] Figure 2 shows how the age of production (A)

[3]This correlation is also compatible with INT, although the causality is reversed. Indeed, from the perspective of this growth scenario, higher connectivity in the global network is caused by earlier learning, not the other way around. Some words end up being highly connected in the global network precisely because they happen to be acquired earlier and, therefore, have a higher chance of accumulating more links over time.

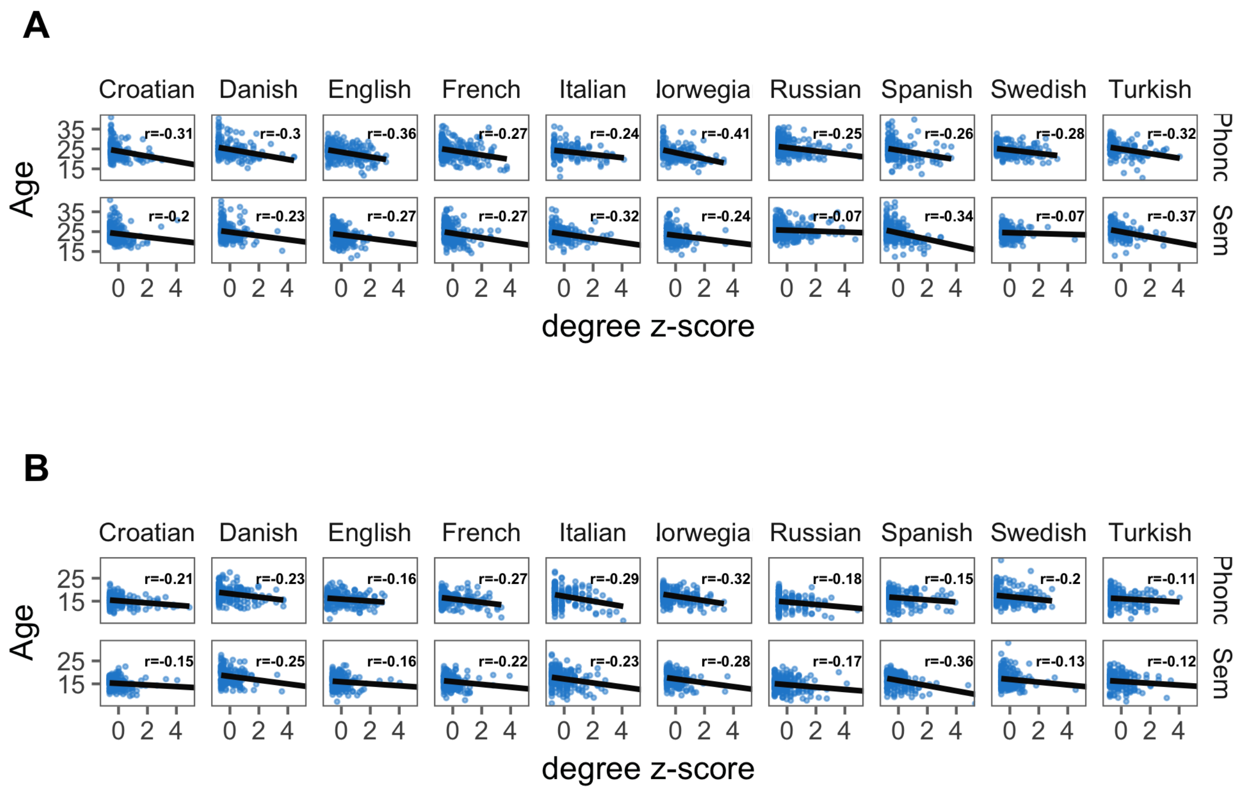

Figure 2. Age of production (A) and comprehension (B) in the global network as predicted by the degree in this network. Results are shown in each language for phonological and semantic networks. Each point is a word, with lines indicating linear model fits.

and comprehension (B) for each word varies as a function of its degree (or indegree for the semantic networks). For ease of visual comparison, the predictor (i.e., the degree) was centered and scaled across languages. The plots show, overall, a negative correlation between the month of acquisition and the degree, indicating that nouns with higher degrees are generally learned earlier. HERE MENTION THE AVERAGE VALUE OF THE CORRELATIONS

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

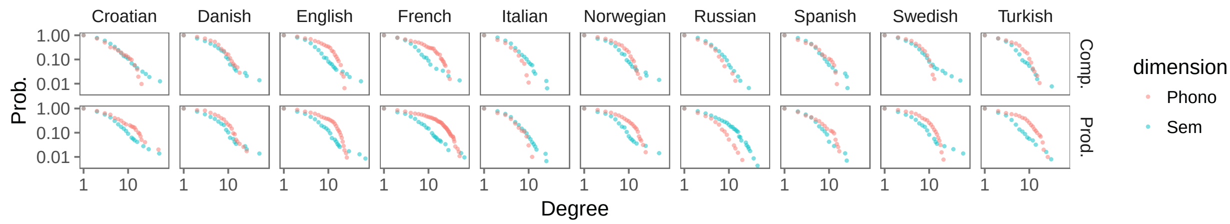

Figure 3. Log-log plot of the cumulative degree distribution function for the global phonological and semantic networks across languages. The figure shows the results for both production and comprehension data. A perfect power-law distribution should appear as a straight line in this graph.

distribution of the form $p(k) \propto \frac{1}{k^{\alpha}}$, Barabasi & Albert, 1999). If the network displays this property, this fact would suggest a INT-like generative process. Conversely, if the degree distribution does not follow a power law, this fact would weaken the case for INT. The log-log plots are shown in Figure 3. We fit a power law to each empirical degree distribution following the procedure outlined in Clauset, Shalizi, and Newman (2009) and using the 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 XX. Overall, we could not reject the null hypothesis of a power-law distribution: the p-value was generally above 0.1. Also talk about the fact that the power law is truncated.

In sum, the static properties of the global network are *a priori* compatible with both

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

INT and EXT. In order to decide between these two developmental scenarios, we need to fit explicit growth models to the data.

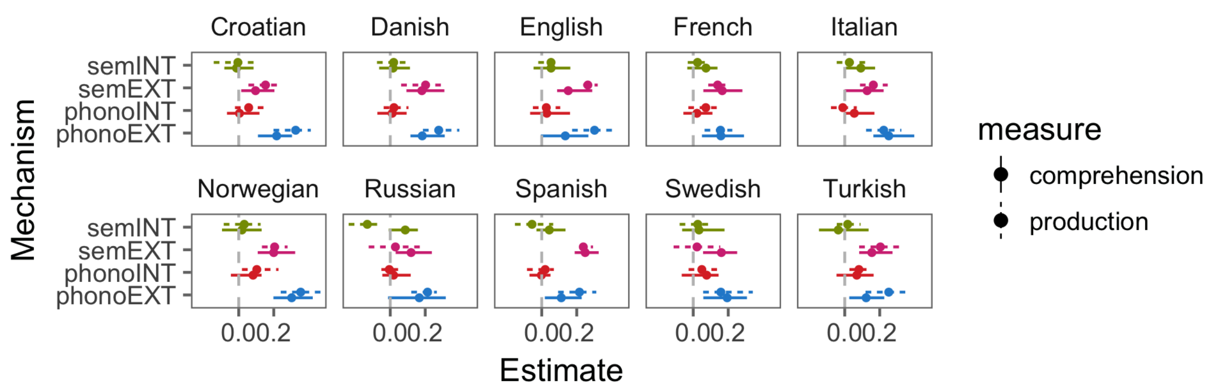

Figure 4. Evaluation of growth scenarios (EXT: externally-driven, INT: internally-driven) for both semantic and phonological networks. Each point represents the mean of the posterior distribution of the growth parameter, with ranges representing 95% credible intervals. Positive values mean that learning 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.

Network growth models

How does each growth scenario predict noun development? To test the network growth scenarios, we fit different growth models to the data. We calculated the probability that a word w_i, with a growth value d_i would enter the lexicon at a given month, using a softmax function:

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

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 growth values d_i are acquired first, and a negative value means that words with lower growth values are acquired first (see Figure 1 for an illustration of how growth values d_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 (N. Goodman & Stuhlmuller, 2014). We defined a uniform prior over β, and at each month, we computed the likelihood function over words that could possibly enter the lexicon at that month, fit to the words that have been learned at that month (using formula 1). Markov Chain Monte Carlo sampling resulted in a posterior distribution over β, which we summarized in Figure 4.

The results replicate Hills et al.'s finding regarding the semantic netowrk in English and prouction data, which is that this network grows by EXT, not by INT. Besides, this finding generalize to comprehension, and holds overall across languages. 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 intial approximation, albeit a posteriori, about the unversisality of the semantic similarity measure. In the phonological domain, the EXT model also fits better the data than INT for both production and comprehension. Similar to the sematic domain, the findings generalize well cross-linguistically.

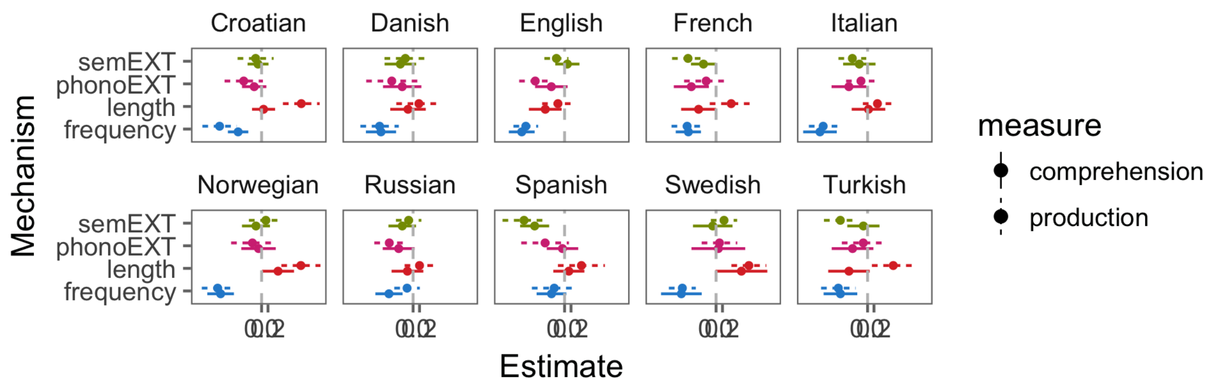

Figure 5. Estimates of the relative contribution of each predictor of AoA in the combined regression model. Results are shown for both production and comprehension data. Ranges indicate 95% confidence intervals. Positive values indicate a positive relationship (e.g. longer words tend to have a higher AoA), while negative values indicate a negative relationship (e.g. words with higher frequency tend to have a lower AoA).

Comparison to other predictors of age of acquisition

We saw that the way semantic and phonological information is structured in the learning environment (i.e., EXT) contributes to noun learning across languages. However, we know that other factors influence learning as well (e.g., Braginsky et al., 2016). Next we investigated how semantic and phonological connectivity interact with two other factors. The first one is word frequency, a well studied factor shown to predict the age of acquisition in a reliable fashion (e.g. J. C. Goodman et al., 2008). The second factor is word length,

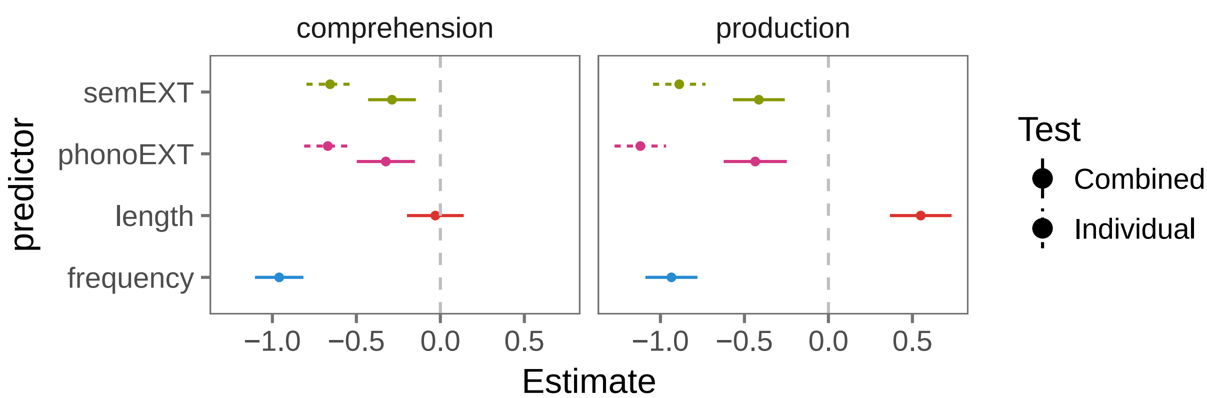

Figure 6. Estimates of the relative contribution of each predictor of AoA in the combined mixed-effects model with language as a random effect. Results are shown for both production and comprehension data. Ranges indicate 95% confidence intervals. Dotted ranges indicate the estimates for the predictor in a separate model that includes only this predictor as a fixed effect.

which correlates with phonological connectivity: Shorter words are more likely to have higher connectivity (Pisoni et al., 1985; Vitevitch and Rodriguez, 2004), and thus it critical to test if phonological connectivity plays a role in learning above and beyond word length.

Since the previous section showed that INT was uninformative, we dropped it from this analysis, keeping only EXT. This simplified the model because we no longer needed to fit growth 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 linear 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. (2016) where unigram counts were derived based on CHILDES corpora in each language (MacWhinney, 2000). For each word, counts included words that shared the same stem (e.g., "cats" counts as "cat"), or words that were synonymous (e.g. "father" counts as "daddy"). Note that these frequency counts are based on transcripts from independent sets of children and represent a general estimate of environmental frequency across children. The use of an independent dataset is warraned because it uses large samples to average out the difference between children (see Goodman et al. 2008).

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 5 shows the coefficient estimate for each predictor in each language for production and comprehension data. Figure 6 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., 2016; J. C. Goodman et al., 2008; B. C. 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., under review). Thus, word length seems to reflect the effects of production constraints rather than than comprehension constraints (i.e., long words are harder to pronounce than short words, but maybe not more difficult to store or access).

As for the factors of interest, i.e., semantic and phonological connectivity, we found cross-linguistic differences. Phonological connectivity contributes to learning in some languages but not in other. In particular, semantic connectivity does not explain variance in English data beyond that explained by phonological connectivity, frequency and length. This contrasts with the original finding in Hills et al. (2009). However, in this previous study, semantic connectivity was not tested in a model that included frequency, length and phonological connectivity as covariates. Another important difference is the number of words tested: Our study uses a larger set of nouns. That said, and despite these cross-linguistic differences, both phonological and semantic connectivity are significant predictors in the combined model.

Discussion

This study provided an analysis of network growth during development. We compared the two major network growth scenarios desribed in the pioneering work of Hills et al. (2009). The first scenario, INT (orignially called Preferentailly Attachement), 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 (orignially called Preferentailly Acquisition) described a model in which the external, global environmental network structure determines learners' growth patterns. Previous word in this line of research (e.g., Hills et al. (2009); Hills et al. (2010); Sailor 2013) has been limited in its scope: It focused on semantic networks, while relying only on words' production as a mesure of their acquisition, and using almost exclusively developmental data from English-learning children. This study test the generality previous findings first by investigting phonological networks together with semantic networks, second by adopting both production and comprehension as measures of learing, and crucially, by comparing the findings across 10 languages.

We found that the orginal 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 scenrio (EXT), not by the internally-driven mechanism (INT). Besides, both semantic and phonological networks contribute to the learning process above and beyond other known predcitors of word learning such as frequency and word length. Second, comprehension-based vocabularies grow in a way similar to production-based vocabularies [HERE ADD MORE NUANCE?]. Finally, the findings were overall similar across the 10 languages we tested. Although we find some cross-lingusitic 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 the limited and partially-overlapping sample of nouns for each language. In fact, both phonological and semantic connectivity are significant predictors 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 means, strikingly, that children are sensitive to highly connected words although they do not initially have access to the full network. This remarks begs the following question: What mechanism could allow children to tease apart highly connected words from low connected words? Besides, why would highly connected words be easier to learn?

One possbility is that children rely on their statistical learning abilities. For example, in the semantic domain, children might be using a mechanism akin to cross-situational learning to pick up the meanings of some words, especially concrete nouns (Smith & Yu, 2008), and such learning would resemble the growth scenario described by EXT. Indeed, since free association is related to contextual co-occurrence (Griffiths et al., Fourtassi and Dupoux, 2013), highly connected words will tend to occur in a variety of speech and referential contexts. This fact makes such a word easier to learn only because it has more referential disambiguating cues across learning contexts, and crucially, even without knowing the entire set of words with witch it occurs (hence the similarity with EXT). This possibility is supported by the finding that nouns' diverstiy of occurrence in child directed speech predicts their age of learing (Hills et al. (2010)).

In the phonological case, network growth according to EXT is also compatible with a scenario whereby children are tracking lwo level statistical patterns, e.g., high probability sound sequences. Indeed, connectivity in the phonological network is inherently correlated with phonotactic probability (Vitevitch, Luce, Pisoni, & Auer, 1999). That is, highly connected words tend to be made of frequent sound sequences. Children are sensitive to local phonotactic regularities (Jusczyk, Luce, & Charles-Luce, 1994) and this sensitivity might lead them to learn higher-probability words more easily (e.g., Storkel, 2001). This explanation is supported by computational simualtions that shows how learning general phonotactics patterns create “well-worn paths” which allow for the representation of several distinct but phonologically neighboring words (Dell et al; 1993; Tavac et al. 2017).

Besides using their own statitistical learning skills, children could also benefit from the the way their caregivers speak. Perhaps the cargivers 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. Investigting 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 infuence by this input (Clark 2007; Hoff and Naigles, 2002; Hurtado et al., 2008).

This work shares a number of limitations with previous studies using similar research strategy and datasets (e.g., Hills et al. 2009; 2010). Chief among these limitations is the fact that the age of word acquisition is computed using different children at different ages (due to the fact that avialable CDI data is mainly cross-sectional). Althoug this measure has proven highly consistent (Fensen et al. 1993), it lead us to focus on studying the learning mechanism of the “average” child. Individual trajectories, however, could show different learning patterns. For example, using longitudinal data Beckage, Smith, and Hills (2011) found differences between typical and late talkers in terms of the semantic network structure. Besides, although our study endorses the externally-driven account of network growth, this does not mean individual children never use some variant of the internally-driven mechanism. For instance, some children develop “islands of expertise”, that is, well organized knowledge about a certain topic (e.g., birds or dinosaurs). This prior knowledge enables these children to learn new related words more easily (e.g., Chi and Koeske, 1983).

To conclude, our work validates previous results in early network development, and suggests that the advantage of highly connected words words may emerge, at least in the early stages of word learning, from the operation of simpler mechanisms in both the semantic and phonological domains. One question for future experimental work is whether 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

Acknowledgements

This work was supported by a post-doctoral grant from the Fyssen Foundation.

Disclosure statement

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

Appendix 1: Power law model

	Kmin	alpha	pValue	dimension	language
1	4.00	2.55	0.88	Sem	Croatian
2	4.00	2.18	0.12	Phono	Croatian
3	4.00	2.38	0.00	Sem	Danish
4	11.00	4.55	0.86	Phono	Danish
5	5.00	2.66	0.13	Sem	English
6	20.00	9.14	0.51	Phono	English
7	8.00	2.81	0.13	Sem	French
8	20.00	3.75	0.11	Phono	French
9	4.00	2.93	0.61	Sem	Italian
10	9.00	9.45	0.78	Phono	Italian
11	5.00	2.88	0.20	Sem	Norwegian
12	15.00	6.28	0.74	Phono	Norwegian
13	24.00	5.61	0.72	Sem	Russian
14	8.00	4.20	0.54	Phono	Russian
15	4.00	2.98	0.46	Sem	Spanish
16	13.00	8.75	0.74	Phono	Spanish
17	4.00	2.49	0.17	Sem	Swedish
18	11.00	4.68	0.10	Phono	Swedish
19	4.00	2.87	0.93	Sem	Turkish
20	8.00	3.26	0.38	Phono	Turkish

Table 2

Results of fitting a power law model to the degree distribution in each model for production data. Kmin is the optimal degree cut-off, alpha is the scaling parameter, and pValue is the probability that quantifies the plausibility of the power law hypothesis. If pValue is close to 1, the power law model cannot be rejected as a plausible fit for the data. If, instead, pValue is small (e.g., $p < 0.05$*) then the null hypothesis of a power law model can be rejected.*

	Kmin	alpha	pValue	dimension	language
1	5.00	2.67	0.90	Sem	Croatian
2	2.00	2.06	0.02	Phono	Croatian
3	4.00	2.39	0.01	Sem	Danish
4	5.00	2.98	0.14	Phono	Danish
5	4.00	2.64	0.77	Sem	English
6	13.00	5.16	0.23	Phono	English
7	4.00	2.63	0.33	Sem	French
8	18.00	5.58	0.34	Phono	French
9	4.00	2.88	0.69	Sem	Italian
10	8.00	10.27	0.91	Phono	Italian
11	5.00	2.87	0.43	Sem	Norwegian
12	13.00	7.65	0.44	Phono	Norwegian
13	8.00	3.91	0.95	Sem	Russian
14	5.00	3.97	0.85	Phono	Russian
15	5.00	3.11	0.55	Sem	Spanish
16	5.00	3.01	0.09	Phono	Spanish
17	5.00	2.81	0.71	Sem	Swedish
18	9.00	6.75	0.10	Phono	Swedish
19	4.00	3.13	0.89	Sem	Turkish
20	9.00	5.73	0.96	Phono	Turkish

Table 3

Results of fitting a power law model to the degree distribution in each model for comprehension data. Kmin is the optimal degree cut-off, alpha is the scaling parameter, and pValue is the probability that quantifies the plausibility of the power law hypothesis. If pValue is close to 1, the power law model cannot be rejected as a plausible fit for the data. If, instead, pValue is small (e.g., $p < 0.05$) then the null hypothesis of a power law model can be rejected.

References

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

Beckage, N. M., Smith, L., & Hills, T. T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLOS ONE*, *6*(5), 1–6.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2016). From uh-oh to tomorrow: Predicting age of acquisition for early words across languages. In *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

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

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: Network distinctiveness predicts age of acquisition. *Cognitive Science*, *41*, 120–140.

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), i–185.

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.

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.

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.

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.

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

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/

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), 12663–12668.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

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.

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.