Word Learning as Network Growth: A Cross-linguistic Analysis

Abdellah Fourtassi
afourtas@stanford.edu
Department of Psychology
Stanford University

Yuan Bian
ybian.uiuc@gmail.com
Department of Psychology
University of Illinois

Michael C. Frank
mcfrank@stanford.edu
Department of Psychology
Stanford University

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: semantic network, phonological network, network growth, mechanism of word learning

Introduction

Over the first year of life, children become sensitive to the phonetic variations that are used to distinguish meanings in their native language (Werker and Tees, 1984). One could imagine that these perceptual skills would be automatically applied to the task of word learning. However, developmental data show that 14 m.o children find it challenging to associate minimally different (but perceptually discriminated) sounds such as "bin" and "din" to different objects (Stager & Werker, 1997).

Several factors can explain this pattern of results. For example it is possible that the task of meaning learning increases cognitive demands on children (compared to a simple perceptual descrimination). In particular, it requires paying attention to both the sounds and the corresponding objects, which may hinder precise encoding in memory of some phonetic details (Stager a& Werker, 1997). Additional difficulty might arise from ambiguous phonological boundaries at this stage of development (e.g., Rost & McMurray, 2009), or from uncertainty about the referential status of the novel word (Fennell & Waxman, 2009).

Regardless of the exact explanation, it is generally accepted that by around 17 m.o, chidlren succeed under the same circumstances (Werker et al. 2002). What could be the mechanism of development? Swingley contrasted a binary-discontinuous scenario to a probabilstic-continuous one. On the discontinuous account, younger children learn a single, underspecified representation of similar words (e.g., "bin"/"din"). Development occurs when children specify this intial coarse reprepsentation and learn two distinct catgories. On the continuous account, distinct represetations are successfully learned from the start, but these representations are encoded with higher uncertainty in youger children, leading to apparent failure in relatively demanding tasks.

Experimental evidence suggest a probabilsitic-continuous, rather than a stage-like develomental scenario. On the one hand, 14-month-olds who typically fail in the original task succeed when learning is probed using a testing method with lower memory demands (Yoshida et al. 1009), and when uncertainty in the emerging representations is mitigated using disambiguating contextual cues (Thiessen, 2007; Dautriche et al., 2016). On the other hand, adults show patterns of learning similar to those shown by 14-month-olds when the task is more challenging and when similarity between words increases (Pajak et al., 2016; White et al., 2013).

Following this experimental studies, the purpose of the current work is to propose a precise probabilsitc model of the Stager and Werker's task, and to explore the extent to which such a model could provide a unified accounts for several documented (as well as new) patterns of learning and development. We first present the model and explain how it can be used to make precise predictions of the target behavioral responses. Second, using new data collected from both preschool children and adults, we show how the probabilistic strucure of the model can explain behavioral patterns both within the same age and across development.

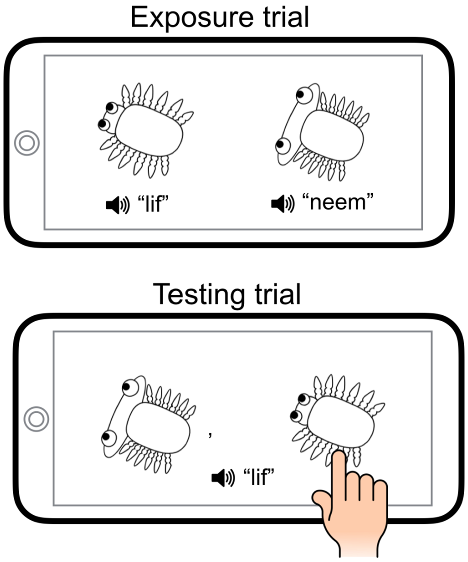

Figure 1: Illustration of the growth scenarios. Filled circles (I1-I4) represent known words (internal), and empty circles (E1 and E2) represent words that have not been learned yet (external). Black lines represent links that are relevant in each growth scenario, and gray lines represent links that are irrelevant.

Model

Task

We model a variant of the task used in Yoshida et al. 2009. In this task, particiants are first exposed to the association between pairs of nonesense words (e.g., "lif"/"neem") and pairs of objects. After this exposure phase, participants perform a series of two-alternative forced choices. In each testing trial, one of the two sounds is uttered (e.g., "lif") and participants choose the corresponding object from the two alternatives.

Probabilistic strcuture

Our model consists of a set of variables describing the general process of spoken word recognition in a referential situation. These variables are related in a way that refelects the simple generative scenario represented graphically in Figure XX. When a speaker utters a sound in the presence of an object, the observer assumes that the object o activated the concept C in the speaker's mind. The concept prompted the cooresponding label L. Finally, the label was physically instantiated by the sound s.

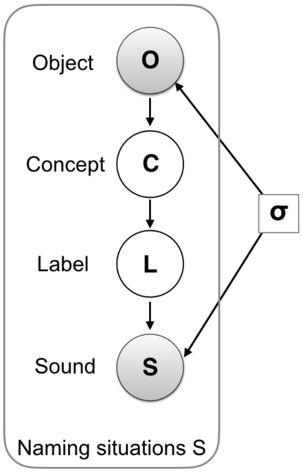

Figure 2: Graphical representation of our model. Circles indicate random variables (shading indicates observed variables). The squares indicates fixed model parameters.

Because of the noisy nature of the representations, the observer can only determine the hidden variables (i.e., the concept C and the label L) in a probabilistic fashion. For simplicity, we assume that the probability of membership of objects and sounds to concepts and labels, respectivel, are normally distributed:

$$p(o|C) \sim \mathcal{N}(\mu_C, \sigma_C^2)$$
$$p(s|L) \sim \mathcal{N}(\mu_L, \sigma_L^2)$$

We assume there to be one-to-one mappings between concepts and labels, and that observers have successfully learned these mappings during the exposure phase:

$$P(L_i|C_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Predictions

During the testing phase (see Figure 1), participants are presented with one target sound $s_T \in \{s_1, s_2\}$ and two possible objects o_1 and o_2. In order to make a choice, we determine which object is more probable under the lable s_T, in other words, we compare the the probabilities $P(o_1|s_T)$ and $P(o_2|s_T)$. The values of these probabilities can be computed by summing over all possible labels:

$$P(o|s) = \sum_{C,L} P(o,C,L|s) \propto \sum_{C,L} P(o,C,L,s)$$

The joint probability $P(o,C,L,s)$ is obtained by factoring the bayesian network in Figure 2:

$$P(o,C,L,s) = P(s|L)P(L|C)P(C|o)P(o)$$

which could be tansformed using Bayes rule into:

$$P(o,C,L,s) = P(s|L)P(L|C)P(o|C)P(C)$$

Finally, assuming that the concept prior is a uniform ditribution, we obtain the following expression, where all conditional dependencies have been defined in the previous secton.

$$P(o|s) = \frac{\sum_{C,L} P(s|L)P(o|C)P(L|C)}{\sum_o \sum_{C,L} P(s|L)P(o|C)P(L|C)} \qquad (1)$$

From the general expression (1) we were able to derive the exact analytical expression of the probability of correct responses in the testing phase of our simplified experimenal setting (Figure 1).

$$P(o_T|s_T) = \frac{1+e^{-(\Delta s^2+\Delta o^2)/2\sigma^2}}{1+e^{-(\Delta s^2+\Delta o^2)/2\sigma^2}+e^{-\Delta s^2/2\sigma^2}+e^{-\Delta o^2/2\sigma^2}} \qquad (2)$$

We simulate the values of this probability as a function of the perceptual distance between the sounds Δs, using the two remining parameters of the model: the visual distance between the semantic referents Δo and the values of the standard deviation of the dsitributions $p(s|L)$ and $p(o|C)$ (which, in this simulation, have similar values, i.e., $\sigma = \sigma_C \approx \sigma_L$). The simulations are shown in Figure 3.

The simulations explain previously documented facts, and make new predictions:

1) For fixed values of Δo and σ, the probability of accurate responses increases as a function of Δs. This pattern accounts for the fact that similar sounds are generally more challenging to learn than different sounds for both chidlren (Stager and Werker, 1997) and adults (Pajak et al. 2016).

2) For fixed values of Δs and Δo, accuracy increases when the noise characterized with σ decreases. This fact may explain development, i.e., youger children have noisier representations (Swingley, 2007; Yoshida et al. 2009), which lead to lower word recognition accuracy, especially for similar sounding words.

3) For fixed values of Δs and σ, accuracy increases when the visual distance between the semantic referents Δo increases.

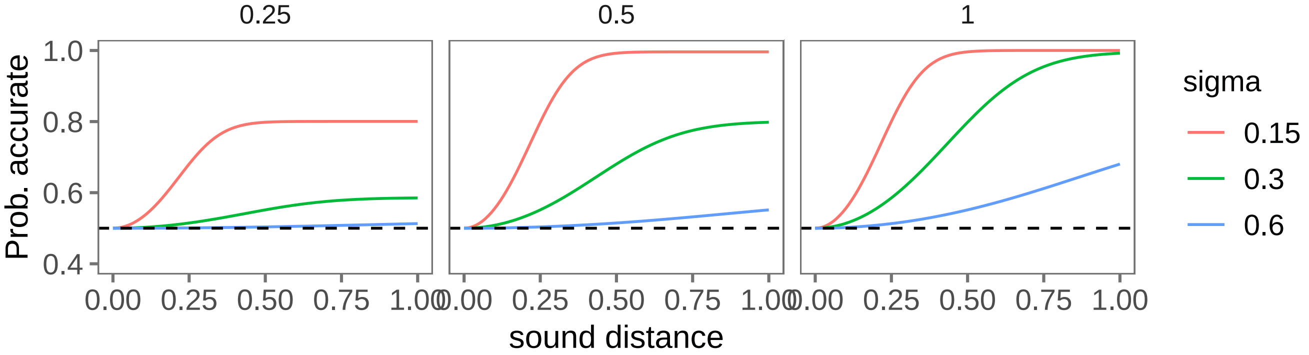

Figure 3: Age of acquisition in the global network as predicted by the degree in this network. Results are shown in each language for phonological and semantic networks. Each point is a word, with lines indicating linear model fits.

This is a new predcition that our model makes. Previous work studied the effect of several bottom-up and top-down properties in disambiguating similar sounding words (see introduction), But no previous study tested the effect of the visual distance between the semantic referents.

To sum, we introduced a model that accounts for some qualitative patterns observed in previous studies, and makes new predictions. In the experiment below, we test whether the model makes accurate *quantitative* predictions by fitting it to new experimental data collected from preschool chidlren and adults.

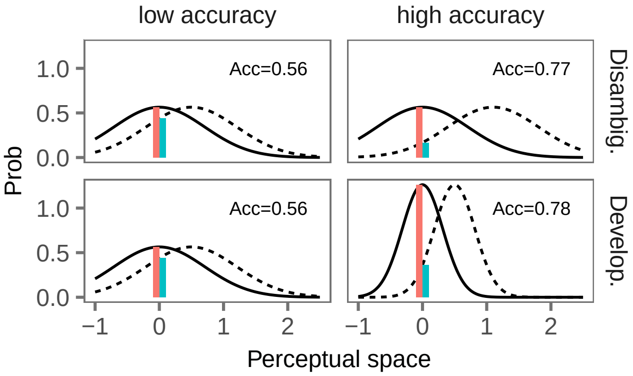

Figure 4: Graphical representation of our model. Circles indicate random variables (shading indicates observed variables). The squares indicates fixed model parameters.

Experiment

In this experiment, we tested participants in the word learning task introduced above (Figure 1). We explored all three parameters of the model. Both the sound similarity (Δs) and object similarity (Δo) were varied simulataneously in a within-subject design. Besides, two age groups (preschool children and adults) were tested on the same task to explore whether development can be characteized with the degree of uncertainty, σ, in the probabilsitic representations.

Methods

Participants We planned to recruit a sample of 60 children ages 4-5 years from the Bing Nursery School on Stanford University's campus. So far, we collected data from N=47 children (mean age= months, F=). An additional 28 children participated but were removed from analyses because they were not above chance on the catch trials (as was specified in the pre-registration[1]). We also collected a planed sample of N=30 adults on Amazon Mechanical Turk. N=2 adult participants were excluded because of low scores on the catch trials (see pre-registration).

Stimuli and similarity rating The sound stimuli were generated using the MBROLA Speech Synthesizer (Dutoit et al., 1996). We generated three kinds of sound pairs which varied in their degree of similarity to English speakers: 1) "different": "lif"/"neem" and "zem"/"doof", 2) "intermediate": "aka"/"ama" and "ada"/"aba", and 3) "similar" non-English minimal pairs: "ada"/"ad^{h}a" (in hindi) and "aʕa"/"aħa" (in arabic).

As for the objects, we used the Dynamic Stimuli javascript library[2] which allowed us to generate objects in four different categories: "tree", "bird", "bug", and "fish". These categories are supposed to be naturally occuring kinds that might be seen on an alien planet. In each category, we generated "different", "intermediate" and "similar" levels of similarity by manipuating a continuous property controling features of the category's shape (e.g, body strech and head fatness).

In a separate survey, $N = 20$ participants recruited on Amazon Mechanical Turk evaluated the similarity of each sound and objcet pair on a 7-point scale. We used the average ratings across participants as qunatitaive measures of sound pairs' distance (Δs) and object pairs' distance (Δo).

Design Each age group saw only two of the three levels of similarity described in the previous sub-section: "different" vs. "intermediate" for preschoolers, and "intermediate" vs. "similar" for adults. The experiment consisted of four conditions which involved, each, one pair of sounds-objects associations. These conditions were constructed by crossing the sound's degree of similarity with the object's degree of similarity leading to a 2x2 factorial

[1] https://osf.io/jrh38/

[2] https://github.com/erindb/stimuli

design in each age group. Besides the 4 conditions, we also tested participants on a fifth catch condition which was similar in its stucture to the other ones, but was used only to select participant who were able to follow the instructions and show minimal learning.

Procedure Preschoolers were asked if they would be willing to play a game on a tablet with the experimenter and were informed that they could stop playing at any time. The experimenter explained that the game consisted in learning some words spoken in an alien planet. The experiment began with two simple examples (not included in the analysis), and in these examples children were given feedback from the experimenter so as to make sure they correctly understood the structure of the task. After the introduction and examples, children were tested in a sequence of five conditions: the four experimental conditions plus the catch condition. In each condition, participants saw a first block of four exposure trials followed by four testing trials, and a second block of two exposure trials (for memory refreshment) follwoed by an additional four testing trials.

In the exposure trials, children saw two objects associated with their corresponding sounds. We presented the first object on the left side of the tablet's screen simultaneously with the corresponding sound. The second sound-object association followed on the other side of the screen after 500ms. For both objects, visual stimuli were present for the duration of the sound clip (800ms). In the testing trials, children saw both objects simultaneousely and heard only one sound. They completed the trial by selecting which of the two objects corresponded to the sound. They responded by touching one of the pictures on the tablet.

The object-sound pairings were randomized across participants, as was the order of the conditions (except for the catch condition which was always placed in the middle of the testing sequence). We also randomized the on-screen position (left vs. right) of the two pictures on each testing trial.

The procedure for preschoolers and adults were identical except that preschoolers were accompagnied by an experimenter and used a tablet, whereas adults used their local computers to complete the experiment online.

Model fitting We fit the analystical expression (equation 2) to the participants' responses in each age group. The values of Δs and Δo were set based on data from the similarity judgment task (described in the stimuli sub-section). We used two models: **model 1** fit only one parameter ($\sigma = \sigma_C = \sigma_L$), and **model 2** fit two parameters ($\sigma_C \neq \sigma_L$). The values of the parameters were derived using weighted least-squares estimates.

Results

First we analyzed the experimental results shown in Figure (XX, left), using a mixed-effects logistic regression with sound and object distances as fixed effects, and with a maximal random effects strcuture (Barr et al. 2013). Results are shown in Table XX. We found a main effect of sound distance on the accuray of learning in both children and adults, thus replicating previous findings. We also found a main effect of object distance, thus confirming the new hypothesis derived from our model.

Table 1: Estimates of predictor coefficients (and their standard errors) by age group in the regression model

	Children	Adults
(Intercept)	0.426* (0.199)	3.114** (1.015)
Sound	0.272** (0.100)	2.320* (0.981)
Object	0.315* (0.137)	2.133* (0.952)
Sound x Object	0.151 (0.097)	1.821 (0.976)
Note:		*p<0.05; **p<0.01; ***p<0.001

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, NSF #1528526, and NSF #1659585.

References

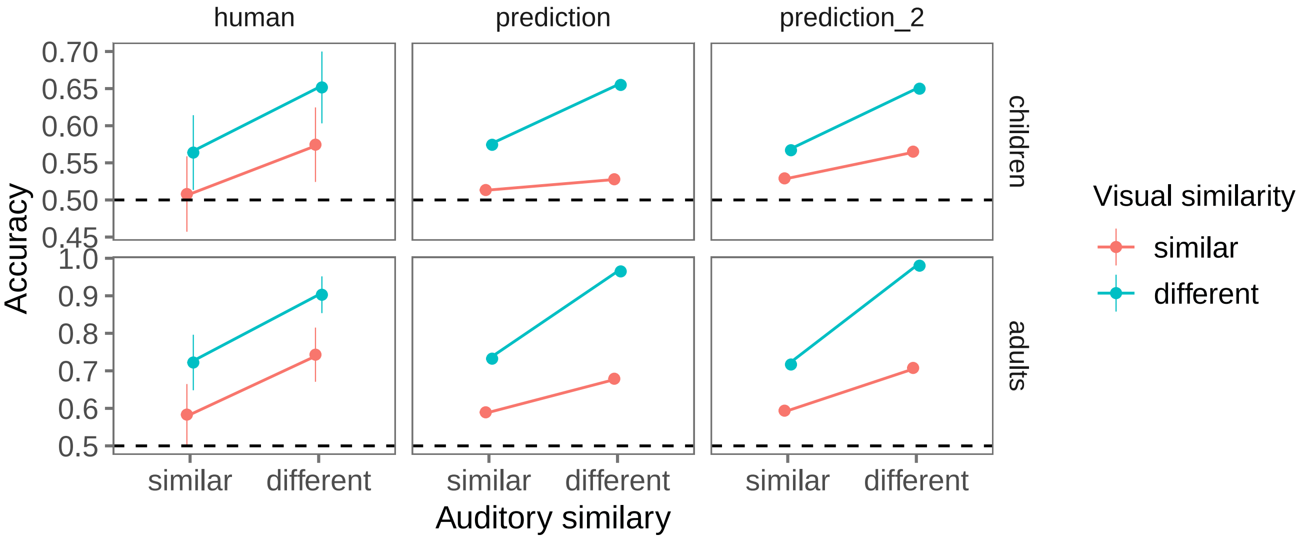

Figure 5: Age of acquisition in the global network as predicted by the degree in this network. Results are shown in each language for phonological and semantic networks. Each point is a word, with lines indicating linear model fits.