UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Developing Semantic Knowledge through Cross-situational Word Learning

Permalink

https://escholarship.org/uc/item/7c351890

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 36(36)

ISSN

1069-7977

Authors

Kachergis, George
Yu, Chen
Shiffrin, Richard

Publication Date

2014

Peer reviewed

Developing Semantic Knowledge through Cross-situational Word Learning

George Kachergis (george.kachergis@gmail.com)
Institute of Psychology / LIBC, Leiden University
Leiden, the Netherlands

Chen Yu and Richard M. Shiffrin
Psychological & Brain Sciences, Indiana University
Indiana, U.S.A.

Abstract

The process of learning a language requires that long-term memory stores the meanings of thousands of words encountered across a variety of situations. These word meanings form a network of associations that, influenced by environmental factors such as word frequency and contextual diversity, cause behavioral effects on measures such as lexical decision and naming times. We investigate the development of recognition priming as a function of explicit knowledge after repeated training and testing on a novel vocabulary. By varying the word frequency and contextual diversity of the training input, and examining learning trajectories as well as semantic knowledge effects, we shed light on which environmental factors most influence performance in language acquisition. Contextual diversity and entropy–the uncertainty about a word's referents–are the two strongest factors predicting primed recognition times, and play a role along with frequency and context familiarity in predicting explicit learning.

Keywords: Language acquisition; cross-situational word learning; semantic knowledge; statistical learning; primed recognition

Introduction

Learning a language is an impressive feat for many reasons, not least of which is the fact that long-term memory stores the meanings of thousands of words–over 12,000 are known at the age of 6 years, while some 60,000 will be known by the end of high school (Bloom, 2000). Short-term laboratory studies have elucidated many of the attentional, learning, and memory mechanisms that contribute to word learning. One important idea has been that of cross-situational learning, which posits that people can learn word meanings by accumulating word-referent co-occurrences from utterances across varying scenes. For example, a learner may hear two novel words w_1 and w_2 while playing with some new toys o_1, o_2, o_3. Later that day, the learner may hear w_3 along with w_1, while attending to o_2 and o_3. If the learner can remember the prior co-occurrence of w_1 and o_2, she may not only strengthen that association, but also learn that w_3 may refer to o_3. Numerous studies have found that adults, children, and even infants can learn nouns via this type of cross-situational training (Yu & Smith, 2007; Smith & Yu, 2008). Adults often perform well despite great uncertainty: 18 pairs to be learned from four word-object pairs at a time, across only five minutes of training. This ability has been modeled using a variety of disparate assumptions about how people store and retrieve possible meanings, ranging from storing mutually-exclusive word-referent hypotheses (Medina, Snedeker, Trueswell, &

Gleitman, 2011) that interact only at storage, to storing memory associations between each word and many referents that compete at storage and retrieval (Yu, 2008; Kachergis, Yu, & Shiffrin, 2012).

Associative models seems more suitable for learning a lexicon, in general, for words are thought not to be stored in isolation, but rather forming a network of associations between both perceptual and verbal contexts, allowing us to define concepts both concrete (e.g., dog) and abstract (e.g., justice). The growth and structure of the semantic network deserves study in its own right, and network analyses have provided insights into how concepts may be added over time both in adults (Steyvers & Tenenbaum, 2005) and developmentally (Beckage, Smith, & Hills, 2011). However, there is little empirical evidence for how quickly these semantic associations form during cross-situational training. The closest empirical work is that of Nelson and Shiffrin (2013), in which participants studied novel Chinese characters during several sessions spanning a few weeks, and showed semantic knowledge with frequency and context effects in a pseudo-lexical decision task. The present study looks for evidence of semantic knowledge developing on an even shorter timescale, conditioned on what explicit lexical knowledge learners manage to acquire during cross-situational training.

Most cross-situational learning experiments provide only one opportunity for measuring knowledge: at the final test, where each word is presented once and an object must be selected. In the present study, we give participants the same block of training trials four times, along with a test between each training block, allowing us to measure individual learning trajectories. In addition to choosing the best object for each word, participants were given a final test of primed pseudo-lexical decision: participants would hear a word, followed immediately by seeing two objects, from which they must quickly choose the studied object. Although this experiment is still quite short in comparison to the typical timescale of real-world language learning, we hope that it will afford us a glimpse at the early development of a lexicon: a system of words learned via accumulated statistical learning. In Experiments 2 and 3, we use a similar design but with different training input. We vary word frequency, which is known to vary widely in the natural language environment (Zipf, 1949), and which correlates with speed-ups in lexical decision times (Scarborough, Cortese, & Scarborough, 1977; Balota & Chumbley, 1984). We also vary contextual diver-

sity (CD), the number of other pairs a given pair appears with across training, which is partially-confounded with frequency and has also been shown to influence word-learning (Kachergis, Yu, & Shiffrin, 2009). In summary, we investigate the impact of various environmental factors on both word-learning trajectories and resulting semantic knowledge.

Experiment 1

Participants were asked to simultaneously learn many word-referent pairs from a series of individually ambiguous training trials using the cross-situational word-learning paradigm. Each training trial was comprised of a display of four novel objects with four spoken pseudowords. With no indication of which word refers to which object, learners had a small chance of guessing the four correct word-referent pairings. However, since a word always appeared on trials with their intended referents, the correct pairings may be learned over the series of trials. In this experiment, we measure learning trajectories by having participants complete four identical training and test blocks, containing the same 18 word-object pairs. Each training block consisted of 27 training trials, allowing each word-object pair to be displayed six times, with four pairs per trial.

Participants

Participants were 66 undergraduates at Indiana University who received course credit for participating.[1]

Stimuli

Each training trial showed an array of four unusual objects (e.g., uncommon tools) while four pseudowords were heard. Eighteen computer-generated pseudowords that are phonotactically-probable in English (e.g., "manu") were spoken by a synthetic monotone female voice. These 18 objects and 18 words were randomly paired for each learner.

Each training block consisted of the same 27 trials. Each training trial began with the appearance of four objects, which remained visible for the entire trial. After 2 seconds of initial silence, each 1 second word was heard (randomly-ordered) followed by two additional seconds of silence, for a total duration of 14 seconds per trial. The training trials were presented in the same order for each block. No pairs were allowed to appear in consecutive trials, constraining the use of working memory to easily infer correct pairings.

Procedure

After each training block, participants were tested for knowledge of word meanings. A single word was played on each test trial, and all 18 referents were displayed (i.e., 18-alternative forced-choice; 18AFC). Participants were instructed to click on the correct referent for the word. Each of the 18 words was presented once, and the test trials were randomly ordered in each block. Learners were not told that

---

[1]For all experiments, none had participated in other cross-situational experiments.

they would be seeing the same trials or even the same stimuli multiple times, but completed four train-test blocks with a short instruction break between each.

A surprise primed 2AFC recognition task was given after the final vocabulary test. Consisting of 54 trials, each trained word was presented twice, along with 18 novel words. On each trial, a single word was heard, immediately followed by two objects displayed side by side. Participants were instructed to quickly indicate which object was one they had been trained on (i.e., 2AFC recognition), regardless of what the word was. Each object appeared three times as a target: once after its referring word (after co-occuring 24 times), once after a spuriously co-occurring word (0-16 times), and once after a novel word (0 co-occurrences).

Results and Discussion

Data from 12 participants were removed because their mean performance on all four vocabulary tests was not above chance (.056 for 18AFC). Data from the remaining 54 participants was analyzed. On average, participants steadily accumulate knowledge, with mean accuracies of 0.18, 0.37, 0.55, and 0.70 at test after successive training blocks (respective SE: .02, .03, .04, .04). Thus, although on average participants only know 3 word-object pairs after the six exposures in block 1, by 24 exposures they have acquired over 10 pairs. After each training block, the proportion of subjects at ceiling on the 18AFC test was 0, .04, .11, and .28.

Mean accuracy on the primed 2AFC recognition task was very high, 0.90 although 3 subjects had performance not significantly above chance (.554), and were thus removed from further analysis. The median RT for the remaining trials was 668ms (sd= 903). The 4.4% of trials with RT slower than 1571ms were trimmed. Accuracy on the remaining trials was .94. We examined correct recognition RTs by prime type, expecting that spurious and novel prime words, which co-occurred rarely or never with the target object, would be slower than matching primes, which appeared 24 times with the target object. We further split out the matching primes into known and unknown matches, corresponding to whether a participant correctly chose the target object for the final 18AFC test of that word. Participants may show a speed-up only when explicit knowledge of the match facilitates choosing the target object, or it may be that implicit knowledge speeds their choices even for unknown matches.

We performed a linear mixed-effects regression to predict the 2,471 correct RTs based on the prime type. Subject was coded as a random effect, and prime type was coded as a main effect, with known matching primes serving as the intercept. The resulting coefficients, displayed in Table 1, show that known matching primes resulted in faster RTs (668 ms) than any of the other prime types, which were 54-73 ms slower. Unknown matches were not much faster than novel or spurious primes, indicating that facilitation depends on a subject's explicit knowledge of the match between the prime word and the target object.

Experiment 1 demonstrated that a short training session–24

Table 1: Regression coefficients for correct RTs in Exp. 1.

Prime Type	Coeff. (ms)	SE	t-value	p-value
Known Match	668	19	35.68	< .001
Novel	67	11	6.29	< .001
Spurious	73	10	6.93	< .001
Unknown Match	54	15	3.58	< .001

repetitions per word-object pair–of cross-situational learning results in facilitated primed recognition only for words that both match the target object and were explicitly known to match that object by the subject at the final 18AFC test. Experiment 2 investigates cross-situational learning trajectories and priming as a function of word frequency.

Experiment 2: Varied Frequency, High Contextual Diversity

Experiment 1 showed that cross-situational learning over a short period–roughly 20 minutes–results in lexical priming, a typical semantic memory effect. Experiment 2 uses a set of training trials with varying stimulus frequency: some pairs appeared 3 times per block, some 6 times, and some 9 times. Word frequency is a factor that is known to vary in natural language environments (Zipf, 1949), and to influence lexical decision RT (Scarborough et al., 1977; Balota & Chumbley, 1984).

Participants

Participants were 50 undergraduates at Indiana University who received course credit for participating.

Stimuli and Procedure

The stimuli and procedure used in Experiment 2 were the same as those used in Experiment 1, except that the composition of the training trials differed. Specifically, the 18 word-object pairs were split into three groups of six that appeared with different frequency across training: 3 times, 6 times, or 9 times per block (12, 24, and 36 appearances across the full training period). The pairs from these different frequency groups (low, medium, and high, respectively) were allowed to co-mingle randomly on trials: e.g., a trial might be composed of two high-frequency pairs, a medium frequency pair, and a low frequency pair, or any combination of frequency groups. That is, like Experiment 1, Experiment 2 has high contextual diversity: any of the 18 pairs may appear with any of the other 17 pairs. As before, four pairs were shown on each trial, and participants were tested for knowledge of the word meanings at the end of each block, with a final primed 2AFC recognition task.

Results and Discussion

Data from six participants were removed because their mean performance on all four vocabulary tests was not above chance (.056 for 18AFC). Data from the remaining 44 participants were analyzed. Figure 1 shows the mean accuracy for
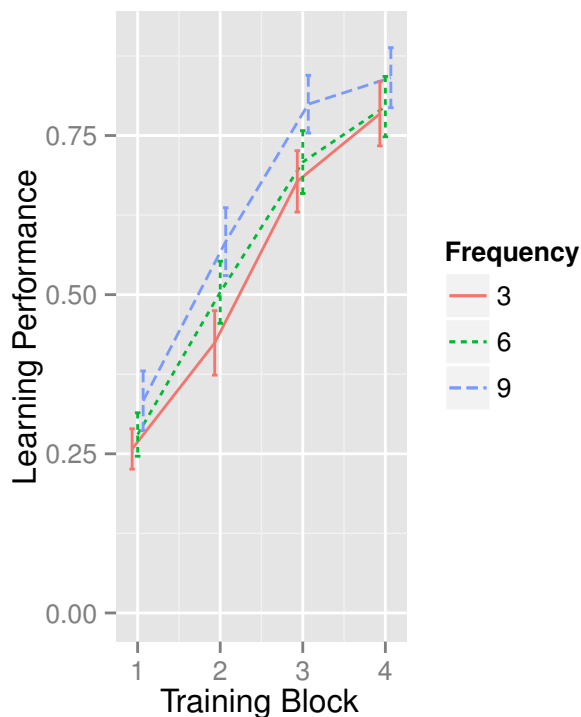

Figure 1: Mean accuracy over training shows very little effect of pair frequency in Experiment 2, likely because low frequency pairs occur with higher frequency pairs.

each frequency group by training block. Surprisingly, lower pair frequency does not seem to hinder performance much. The overall mean accuracy after each training block in Experiment 2 was: .29, .50, .73, and .81, somewhat higher than each block in Experiment 1. After each training block, the proportion of subjects at ceiling on the 18AFC test was 0, .07, .25, and .45, again higher than in Experiment 1.

The median RT for the primed recognition task was 689ms (sd=843). The 6.5% of trials with RT slower than 1532ms were trimmed. Accuracy on the remaining trials was .910. Mean accuracy on the primed recognition task was as high as in Experiment 1, 0.90 although 3 subjects again had performance not significantly above chance (.554), and were thus removed from further analysis. The median RT for the remaining trials was 676ms (sd=845). The 5.8% of trials with RT slower than 1521ms were trimmed. Accuracy on the remaining trials was 0.95.

As in Experiment 1, we examined correct recognition RTs by prime type, expecting that spurious, novel, and unknown matching prime words, would be slower than known matching primes. We again performed a linear mixed-effects regression to predict the 1,975 correct RTs based on prime type, with subject as a random effect and prime type as a main effect (known matching primes as intercept). We also included the frequency of prime-target co-occurrences across all of training as a predictor (0 for novel primes, 0 or 8 for spu-

rious, and 12, 24, or 36 for matching primes). The resulting coefficients, displayed in Table 2, show that known matching primes resulted in significantly faster RTs (619 ms) than novel or spurious primes (99 and 127 ms slower), but only marginally faster than unknown matches (39 ms), unlike Experiment 1. In addition, greater prime-target co-occurrences actually produced a small but significant slow-down (2.6 ms per co-occurrence). These intriguing findings are explored in-depth in a final meta-analysis across all experiments that considers additional environmental factors.

Table 2: Regression coefficients for correct RTs in Exp. 2.

Prime Type	Coeff. (ms)	SE	t-value	p-value
Known Match	619	27	22.76	$< .001$
Novel	99	21	4.65	$< .001$
Spurious	127	21	6.15	$< .001$
Unknown Match	39	21	1.83	$= .07$
Co-occurrences	2.6	0.73	3.58	< 0.001

In summary, Experiment 2 investigated whether varying word and object frequency–which vary in naturalistic situations–would influence the priming effects found only for matching words in Experiment 1. Not only did the priming results not change–showing no infuence of frequency–but even the learning trajectories did not show much influence of frequency. We suggest this may be an effect of contextual diversity: since the pairs of different frequency were allowed to appear with each other, once high frequency pairs are known, learners can shift attention to the low frequency pairs and bootstrap learning. Experiment 3 limits the contextual diversity of the different frequency groups to see whether the priming results are affected.

Experiment 3: Varied Frequency, Low Contextual Diversity – 3 Pairs/Trial

After Experiment 2 found no effect of word frequency on priming–nor even much effect on learning, Experiment 3 uses training trials with limited contextual diversity to determine whether pure frequency affect priming and learning.

Participants

Participants were 70 undergraduates at Indiana University who received course credit for participating.

Stimuli and Procedure

The stimuli and procedure used in Experiment 3 were the same as those used in Experiments 1 and 2, except that the composition of the training trials differed. First, only three pairs were shown per trial, resulting in shorter 11-second trials, and there were 36 trials per block. As in Experiment 2, there were three frequency groups: low (3 times per block), medium (6 times), and high (9 times). However, pairs within these groups were only allowed to appear on trials with pairs of the same frequency: e.g., a trial will have three pairs of
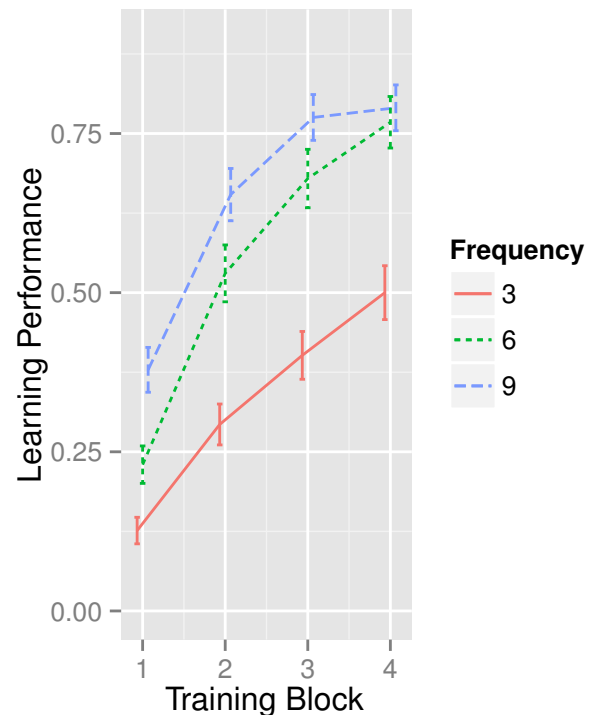

Figure 2: Mean accuracy by frequency group over training shows an advantage for high frequency pairs in Experiment 3. Low frequency pairs cannot be inferred from appearing with more common ones, since they do not co-occur.

either low, medium, or high frequency. Thus, a pair's contextual diversity was limited to within frequency group (the other 5 pairs), and will be lower than in Experiments 1 and 2.

Results and Discussion

Data from four participants were removed because their mean performance on all four vocabulary tests was not above chance (.056 for 18AFC). Data from the remaining 64 participants were analyzed. Figure 2 shows the mean accuracy for each pair frequency group by training block. Unlike Experiment 2, which mixed pairs of differing frequency on the same trials and found only a small frequency effect on performance, keeping the frequency groups separate in Experiment 3 resulted in a larger frequency effect, especially on the low frequency group which suffered most. The overall mean accuracy after each training block in Experiment 3 was: .24, .49, .62, and .69, similar to Experiment 2 in the first two blocks, but somewhat lower in the final two. After each training block, the proportion of subjects at ceiling on the 18AFC test was 0, .03, .09, and .16; nearly a third the proportion at ceiling in the final block of Experiment 2.

Mean accuracy on the primed recognition task was very high, 0.94, although 3 subjects had performance not significantly above chance (.554), and were thus removed from further analysis. The median RT for the remaining trials was

694ms (sd= 681). The 6.5% of trials with RT >1374ms were trimmed. Accuracy on the remaining trials was 0.97.

As in Experiments 1 and 2, we performed a linear mixed-effects regression to predict the 3,096 correct RTs based on prime type, with subject as a random effect and prime type as a main effect. We again included the frequency of prime-target co-occurrences across all of training as a predictor (0 for novel primes, 0 or 8 for spurious, and 12, 24, or 36 for matching primes). The resulting coefficients, displayed in Table 3, show that known matching primes resulted in significantly faster RTs (689 ms) than novel, spurious, or unknown matching primes (46, 49, and 58 ms slower), in results that qualitatively match Experiment 1's. Moreover, prime-target co-occurrences produced no significant effect on correct RT, in contrast to Experiment 2–and despite apparent differences in learning performance at different frequency. These findings are explored more in-depth in the following section.

Table 3: Regression coefficients for correct RTs in Exp. 3.

Prime Type	Coeff. (ms)	SE	t-value	p-value
Known Match	689	21	33.05	< .001
Novel	46	16	2.89	< .01
Spurious	49	16	3.14	< .01
Unknown Match	58	15	4.72	< .001
Co-occurrences	-0.16	0.53	0.20	0.76

Meta-analysis

Word-learning Performance

To better understand what predicts success in cross-situational learning we examined several potentially influential factors across all three experiments, measured for each word in the training trials. We measured word frequency (ranging from 3, 6, and 9 after block 1 up to 12, 24, and 36 after block 4) and contextual diversity (CD; number of distinct objects a word is heard with during training, range: 4-14), which have previously been shown to influence cross-situational word learning (Kachergis et al., 2009). We also measured age of exposure (AE; the trial at which a word first appears, range: 1-13) and context familiarity (CF; the mean familiarity of words heard with a given word, where familiarity is the number of prior appearances, range: 1.44-5.72), which have been suggested to be more important than CD at determining performance (Fazly, Ahmadi-Fakhr, Alishahi, & Stevenson, 2010). Finally, the entropy of each word w's possible associates (i.e., $\sum_o p(o,w)log(p(o,w))$) was included to measure the uncertainty (i.e., dispersion of belief) about w's meaning. Entropy is an important component of a recent cross-situational learning model which assumes learners are biased to attend to uncertain stimuli (Kachergis et al., 2012).

Using these factors to predict performance on 11,808 test trials across training, we fit a logistic regression model using the lme4 package in R (Bates & Maechler, 2010; R Core Team, 2013). For comparison, a null model with only an intercept has AIC = 16,367[2]. We fit a model with an intercept and all five factors (frequency, CD, AE, CF, and entropy) as main effects with no interactions. This model's coefficients[3], displayed in Table 4, (AIC = 14,339), show that all factors were significant except AE. Greater frequency and CF had the expected positive effect on performance, but CD and entropy had effects that may at first glance appear surprising. Lower CD moderately increased learning, perhaps because appearing with fewer pairs results in fewer competing associations. Yet increased entropy (i.e., uncertainty), which may serve as an attentional cue (Kachergis et al., 2012), predicted higher learning. Yet entropy and CD are correlated and trade off: Fitting a model without entropy yields a smaller CD effect (-.03, $p < .001$), but fits slightly better (AIC = 14,371) than a model without CD (AIC = 14,379), in which entropy is estimated to have a significant negative effect (-.19, $p < .001$). The interactions cannot be explored in this space, but frequency, entropy, CF, and CD all contribute to predict cross-situational word learning performance. Now we examine which of these factors predict target recognition for matching primes.

Table 4: Regression coefficients for learning performance.

	Coefficient	SE	Wald's z	p-value
Inter.	-4.69	0.572	-8.20	< .001
Freq.	0.06	0.007	9.32	< .001
Ent.	1.93	0.328	5.87	< .001
AE	0.01	0.009	0.99	.32
CD	-0.30	0.047	-6.51	< .001
CF	0.09	0.011	7.50	< .001

Semantic Knowledge

We did a mixed-effects linear regression to predict the correct RTs for matching primes–known and unknown–according to the same environmental training factors (AE, CD, CF, frequency, and entropy). Additionally, instead of only using the final 18AFC's test as a predictor for correct RT, we used the subjects' average accuracy (Acc.) on that word, across all four tests. Shown in Table 5, mean accuracy, CD, and entropy were significant predictors of correct RT for matching primes, while CF and AE were not, and frequency was marginally significant. Higher 18AFC accuracy sped recognition of that object, as did greater entropy of the prime's associates. In contrast, higher CD or frequency produced slower RTs.

Discussion

We have presented data that suggests implicit knowledge can develop during the course of a single cross-situational learning experiment. The differences in all three experiments are

---

[2]Akaike's information criterion (AIC) is used for model selection; lower is better.

[3]The coefficient estimates a predictor's effect size in terms of log odds ratio.

Table 5: Regression results for correct matching prime RT.

	Coefficient	SE	t-value	p-value
Inter.	966.1	127	7.58	< .001
Acc.	-123.3	13.30	-9.27	< .001
Freq.	1.6	0.93	1.69	=0.09
CF	1.8	6.16	0.29	=.77
AE	0.1	1.72	0.08	=.94
CD	19.1	9.30	2.06	< .05
Ent.	-144.2	69.21	-2.08	< .05

consistent: when primed with the matching word, people are faster to recognize the old object than when primed with either a novel word or with a spurious old word. Additionally, we measured several environmental factors–word frequency, context familiarity (CF), contextual diversity (CD), entropy of associates, and age of exposure (AE)–at the item level for each experiment, and performed a meta-analysis of both the explicit learning trajectories and of the correct primed recognition RTs for matching primes. This analysis found that although greater frequency, CF, and entropy improve word learning, only greater entropy improves recognition RT–frequency marginally slows RT, and CF has no effect. Increased CD harms explicit word learning, and further slows recognition responses. Overall, we have shown that a variety of environmental factors influence word learning in diverse ways, and that their impact can be found not only in explicit word learning, but also in primed recognition, which is akin to the lexical decision task, assessing semantic knowledge.

Moreover, whereas word frequency is a metric of a single word, CD and entropy are metrics of the distribution of the lexicon with a word. Their significance in predicting both learning and semantic knowledge effects indicates the importance of these distributional statistics to learning, and suggests that people's learning mechanisms are sensitive to these statistics. This lends support to associative models that noisily store many word-object co-occurrences (Kachergis et al., 2012), in contrast to models that store single hypotheses with no contextual information (Medina et al., 2011). Indeed, without storing something of a word's context, how can such a system learn the rich network of the human lexicon? A recent modeling study of naturalistic parent-child interactions showed that hybrid models which learn associations not only between words and object but between words and other words provide a significant learning advantage (Kievit-Kylar, Kachergis, & Jones, 2013). We suggest that the next generation of cross-situational learning models should incorporate such mechanisms, and look to semantic knowledge effects for further support.

References

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions good measures of lexical acces? the role of word frequency in the neglected decision stage. *Journal: Humany Percep-* *tion and Performance*, *10*, 340–357.

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed-effects models using s4 classes. r package version 1.0-5.* Available from http://CRAN.R-project.org/package=lme4

Beckage, N., Smith, L. B., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS One*, *6*(5).

Bloom, P. (2000). *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Fazly, A., Ahmadi-Fakhr, F., Alishahi, A., & Stevenson, S. (2010). Cross-situational learning of low frequency words: The role of context familiarity and age of exposure. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (p. 2362-2367). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2009). Frequency and contextual diversity effects in cross-situational word learning. In N. A. Taatgen & H. van Rijn (Eds.), *Proc. of the 31st annual meeting of the cognitive science society* (p. 2220-2225). Austin, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word–referent mappings. *Psychonomic Bull. & Review*, *19*(2), 317–324.

Kievit-Kylar, B., Kachergis, G., & Jones, M. N. (2013). Naturalistic word-concept pair learning with semantic spaces. In *Proceedings of the 35th annual conference of the cognitive science society*.

Medina, T., Snedeker, J., Trueswell, J., & Gleitman, L. (2011, May). How words can and cannot be learned by observation. *PNAS*, 1–6.

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, *120*(2), 356–394.

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria. Available from http://www.R-project.org/

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(1–17).

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

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

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language Learning and Development*, *4*(1), 32–62.

Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.

Zipf, G. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.