

# The Interplay of Multiple Mechanisms in Word Learning

Judith Koehne & Matthew W. Crocker

Department of Computational Linguistics

Saarland University

Saarbrücken, Germany

{judith, crocker}@coli.uni-saarland.de

## Abstract

Word learning in adults succeeds with the help of various mechanisms and is based on multi-modal information sources. The complex interplay of these different cues, however, has rarely been studied. We present two experiments investigating how cross-situational word learning (CSWL) and learning based on sentence-level constraints (SLCL) interact. Our results reveal that SLCL reduces the impact of CSWL when cues are in conflict (Experiment 1) and even blocks statistical sensitivity when cues are independently applicable (Experiment 2). We suggest that the probabilistic nature of CSWL and the more deterministic cues offered by SLCL may underlie this behavior.

**Keywords:** Language learning; cross-situational word learning; sentence-level constraints;

## Introduction

Disadvantaged as they may be in some respects, adult language learners benefit from two natural characteristics. Firstly, they are sensitive to the informativeness of various kinds of available sources such as regularities regarding the linguistic input and its context, the visual environment, and social cues. Secondly, and more so than children, they can constantly connect these multi-modal perceptions and cues with their rich prior knowledge about both language structures and the world. Results from a number of studies reveal that language novices track co-occurrences between unknown spoken words and visual referents across situations (*cross-situational word learning*, CSWL, e.g., Quine, 1960; Yu and Smith, 2007). Additionally, the linguistic context can constrain word meaning, for instance via the relation between a verb and its arguments (e.g. subject and direct object): On the one hand, the arguments define the rough semantic category of the verb (*syntactic bootstrapping*, Landau and Gleitman, 1985; Lee and Naigles, 2008); on the other hand, the verb's semantic restrictions can narrow down the category of the direct-object noun learners need to consider (Koehne and Crocker, 2010). Adults rapidly integrate their spontaneous intuitions about plausible relations on-line, for example to anticipate objects (Altman and Kamide, 1999), and, as language learners, frequently make use of inferencing strategies when words are unknown (Field, 2004).

Only few studies have addressed the interplay of different word-learning mechanisms relying on these kinds of multi-modal cues. Gillette (1999) found that combined linguistic context (verb frame and lexical information) and scene information can result in better verb learning than only one of these cues. Koehne and Crocker (2010) present evidence for the boost of CSWL by learning based on supporting sentence-

level constraints (SLCL), in particular the combination of verbal restrictions (of verbs such as *to eat*), visual scene, and prior language-world knowledge. These studies still make idealizations, however. Different cues are, firstly, fully in accordance with one another and, secondly, available simultaneously. In realistic learning scenarios, this is not necessarily the case: Learning cues are imperfect, information is frequently ambiguous and sometimes conflicting. It is therefore important to examine how helpful different cues are when in conflict, how they influence each other's use, and which are prioritized over others. Moreover, the moment in which potentially helpful sources are available is often not the same moment in which this information can be used. While this difficulty is rarely taken into account, one exception is the recent study by Arunachalam and Waxman (2010), which suggests that syntactic bootstrapping still works when verb information is not co-present with the visual referent.

Studying the different possible scenarios of interacting learning mechanisms potentially also provides information about the underlying nature of these mechanisms, an issue that has rarely been discussed within the empirical word-learning research (but see Yu and Ballard, 2007, Frank, Goodman, and Tenenbaum, 2009, and Alishahi and Fazly, 2010 for formalizations based on computational models). To conduct CSWL, it is necessary for learners to consider different mappings between unknown words and potential visual referents in parallel. That means that this way of learning is non-direct in that more than one hypothesis needs to be maintained, at least until the first theoretically disambiguating situation. There is some evidence that conducting CSWL works probabilistic and in parallel (Yurovsky, Fricker, Yu, and Smith, 2010). In particular, learners not only seem to store more than one mapping between an unknown word on the one hand and potential referents on the other hand, but are also sensitive to fine-grained differences in co-occurrence frequencies (Vouloumanos, 2008). Gaze, gesture, or sentence-level constraints, on the contrary, potentially offer a more deterministic way of learning because these cues are often unambiguous and therefore directly and immediately helpful. We therefore expect such cues to be more reliable for the learner than cross-situational co-occurrence statistics. Furthermore, SLCL as investigated by Koehne and Crocker (2010), additionally appears to exploit semantic category information (e.g. *to eat* selects for objects of the category food).

Due to these differences in the nature of CSWL and SLCL, we hypothesize that SLCL may modulate the use of CSWL when both cues are in conflict (Experiment 1) or indepen-

dently applicable (Experiment 2). We further hypothesize that SLCL still helps noun learning when restrictive verbs and matching visual referents are not co-present, that is, when verb information has to be used across trials (as in Arunachalam and Waxman, 2010; Experiment 1). Finally, we explore whether the nature of the emerging word meanings differ depending on learning strategy (Experiment 2): We hypothesize that while CSWL users are sensitive to fine-grained statistical differences in co-occurrences of unknown words and potential referents, SLCL users are more likely to associate category-based features of potential referents.

## Experiment 1

### Methods

**Participants** 28 German native speakers took part in Experiment 1, four of which had to be excluded due to unsuccessful verb learning. Data from 24 learners was analyzed (mean age 24, 20 females).

**Design, Materials & Procedure** The experiment sought to teach participants a miniature semi-natural language (modified Indonesian) consisting of two restrictive verbs ('eat', 'sew'), two non-restrictive verbs ('take', 'point at'), twelve nouns ('man', 'woman', ten object names), and the article *si*. It comprised the following main stages: verb learning, noun-learning Block 1, Vocabulary Test 1, noun-learning Block 2, Vocabulary Test 2, noun-learning Block 3, Vocabulary Test 3.

In Phase 1, participants familiarized themselves with the four verbs. First, they watched action animations while hearing spoken verbs. Then, pictures of the four actions were visible at the same time (the last position of the animations), one verb was played, and participants were requested to click onto the action matching the verb. Finally, animations were presented silently and participants named the actions themselves. We were not interested in the process of verb acquisition itself but participant's verb knowledge was a necessary prerequisite to investigate the effect of verbal constraints on noun learning.

In the three noun-learning phases, participants were exposed to pairs of static scenes and spoken subject-verb-object (SVO) sentences (sentence start 1s after picture). Sentences consisted of unknown nouns and the just-learned verbs (e.g. *Si laki tambamema si sonis*, 'The man takes the SONIS'). Scenes generally depicted inanimate objects (referents of the nouns and distractors) as well as agent characters and some background. Learners' task was to understand the sentences and learn the ten object names. There were 60 trials, each of the ten novel nouns was presented six times.

Each noun, importantly, had two potential meanings (i.e. referents). One of the two meanings for each noun was supported by CSWL: The co-occurrence of the noun with that object was 83% (high-frequency object, 'socks' in Table 1). The other meaning was less supported by CSWL (co-occurrence only 50%, low-frequency object, 'corn'). Objects other than the high-frequency object and the low-frequency object, the distractors, all co-occurred only once with one

Table 1: Example trials for the noun *bintang*, Exp. 4

trial	verb	depicted objects
<i>Condition Non-restrictive</i>		
1	<i>take</i>	socks (83%), corn (50%), dress (17%)
2	<i>point at</i>	socks (83%), corn (50%), top (17%)
3	<i>point at</i>	socks (83%), corn (50%), pizza (17%)
4	<i>take</i>	socks (83%), jacket (17%), jumper (17%)
5	<i>take</i>	socks (83%), skirt (17%)
6	<i>take</i>	none
<i>Condition Restrictive</i>		
1	<i>take</i>	socks (83%), corn (50%), dress (17%)
2	<i>point at</i>	socks (83%), corn (50%), top (17%)
3	<i>point at</i>	socks (83%), corn (50%), pizza (17%)
4	<i>eat</i>	socks (83%), jacket (17%), jumper (17%)
5	<i>eat</i>	socks (83%), skirt (17%)
6	<i>eat</i>	none

noun (=17% of the six presentations of the noun).

Additionally, each noun was in one of two conditions: In Condition *R* (*restrictive verb*), it occurred with a restrictive verb in half of the trials. In Condition *N* (*non-restrictive verb*), it always occurred with a non-restrictive verb. In Condition *R*, the meaning which was less supported by CSWL (low-frequency object), however, was supported by SLCL, that is, by the restrictive verb. That means that, while in Condition *N*, there was one clearly supported meaning (the high-frequency object), in Condition *R*, one meaning was supported by CSWL (the high-frequency object), and one meaning was supported by SLCL (the low-frequency object). Table 1 illustrates the way a noun (*bintang*) was presented in both conditions (i.e. with which verb and with which objects in the scene): In three of six trials, both the high-frequency candidate and the low-frequency candidate were depicted, in two trials the high-frequency candidate but not the low-frequency candidate was visible, and in one trial neither appeared on the scene. Importantly, in Condition *R*, restrictive verbs were used only in that half of the trials in which the scene did not include the low-frequency (= SLCL-supported) referent. That means that restrictive verbs and referents matching the verb's semantic category were never co-present and verb information had to be memorized across adjacent trials. The presentation of trials was pseudo-randomized. Two object nouns were presented in Block 1 (12 trials), four in Block 2 (24 trials), and four in Block 3 (24 trials). Pictures were counter-balanced regarding absolute and relative positions.

In the vocabulary-test phases, learners were presented all 20 objects depicted on the screen and heard one spoken noun (10 trials). They were asked to click onto the referent that they believed matched the noun. After decision, they indicated on a rating scale how confident they were about their

choice (1(very unsure)-9(very sure)). The experiment lasted about 40 minutes.

**Predictions** We expected learners to be able to use verb information across trials, as demonstrated by the finding of Arunachalam and Waxman (2010). Additionally, we predicted a clear preference in Condition N for the high-frequency object to be selected in the vocabulary test. For Condition R, however, we expected verb constraints to modulate cross-situational statistical learning, with more learners preferring the low-frequency target than in Condition N.

### Data Analysis, Results, & Discussion

Performance in noun learning (= learning either the low-frequency or the high-frequency meaning) was clearly better than chance (10%): 87.5% for Condition N ( $t(23) = 24.665, p < .001$ ), and 80.8% in Condition R ( $t(23) = 20.206, p < .001$ ). Importantly, there was a main effect of condition for the chosen meaning ( $\chi(1) = 59.30, p < .001$ ): In Condition N, learners chose the high-frequency meaning 97% of the times and the low-frequency meaning only 3%. In Condition R, however, both meanings were chosen about equally often (high-frequency candidate: 48%; low-frequency candidate: 52%; see Figure 1). This also confirms that learners were able to use verb information across trials. The average confidence rating was 6.9 and there was no difference between conditions (6.8 in Condition N and 7.0 in Condition R).

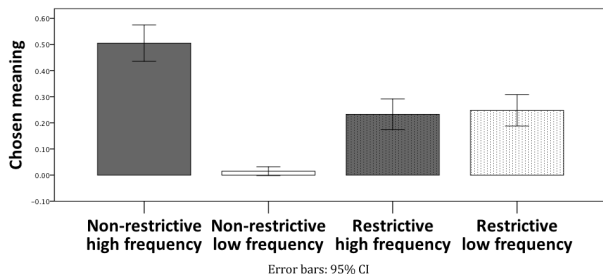


Figure 1: Chosen meanings Experiment 1

Learners' decisions in the vocabulary test reveal a clear difference of condition: While the high-frequency object was unambiguously favored in Condition N, both the high-frequency object (supported by CSWL) and the low-frequency object (supported by SLCL) were chosen equally often in Condition R. This shows that SLCL (verbal constraints) CSWL (co-occurrence frequencies) had a very similar impact on vocabulary decision, with verb information overriding cross-situational statistical information in 50% of the cases. Further, our findings demonstrate how learners make use of verbal restrictions across trials (in accordance with the results of Arunachalam and Waxman, 2010).

## Experiment 2

The goal of Experiment 2 was to further investigate the interplay of CSWL and SLCL when information is not in conflict

but independently applicable: That is, neither contrary as in Experiment 1 nor complementary as in Koehne and Crocker (2010) but redundantly co-present. Specifically, we aimed to investigate the underlying mechanisms of CSWL and SLCL (parallel vs. deterministic). Further, we examined whether whether SLCL enhances learner's sensitivity for category associations.

### Methods

**Participants** 29 German native speakers took part in Experiment 2, five of which had to be excluded. Data from the remaining 24 learners (19 females, mean age 24) was entered into analyses.

**Materials & Procedure** The experimental materials and procedure were similar to those in Experiment 1. The language comprised 18 nouns (the two character names and 16 object names), the same four verbs and the same article as in Experiment 1.

The experiment consisted of the following phases: verb-learn training, noun learning Block 1, Vocabulary Test 1, noun learning Block 2, Vocabulary Test 2.

Participants were introduced into the experiment, verbs were trained and tested exactly as in Experiment 1. Next, learners were introduced into the noun-learning phase. Noun learning consisted of 96 scene-sentence pairs, six presentations per object name. Each noun, again, had two potential meanings (=visual referents), one co-occurred with the noun in 83% of the trials (high-frequency object) and one co-occurred in 50% of the trials (low-frequency object). Nouns were also in one of two conditions: In Condition N (*non-restrictive*), they always occurred with a non-restrictive verb. In Condition R (*restrictive*), they occurred with a restrictive verb in 83% of their presentations (five of six trials). Importantly, in these restrictive trials, there was only one object depicted that matched the verbal restrictions. Unlike in Experiment 1, CSWL (co-occurrence frequencies) and SLCL (verb restrictions) in Condition R supported the *same* meaning: The high-frequency meaning was also supported by the verb. That means that, in Condition R, there was a double cue for learning the high-frequency meaning. There were always four objects on the scene (and sometimes an agent character). Crucially, in three of six trials, both the high-frequency object and the low-frequency object were depicted. In two of six trials, the high-frequency object but not the low-frequency object was included. In one of six trials, none of both referents was on the picture. Distractors, again, all co-occurred only once with one noun (= 17%). The presentation of trials was pseudo-randomized and pictures were counterbalanced as in Experiment 1.

In the vocabulary test, learners heard a noun and were asked to decide for one of four visual objects by clicking on it. There were two different test types. In Test Type 1, the high-frequency object, the low-frequency object, and two distractors were depicted. In Test Type 2, the low-frequency object, two distractors and a *category associate* (CA) were

depicted. The category associate was an object which shared the semantic category with the missing high-frequency object. Each forced choice was followed by a confidence rating, as in Experiment 1. There were 24 test trials (12 per test type), each object name was used twice, once in each test type, respectively. Eight object names each were trained and tested in Block 1 and eight in Block 2.

**Predictions** We expected to find differences between conditions and test types. For Test Type 1, we predicted that learners choose the high-frequency candidate more often than the other objects in both conditions, however, with a clearer dominance in Condition R than Condition N: While in Condition R both SLCL and CSWL support the high-frequency meaning, in Condition N, only CSWL can be used. For test trials of Test Type 2, we predicted a tendency for learners to choose the low-frequency meaning in Condition N because it is statistically the most plausible alternative to the high-frequency meaning and we hypothesized CSWL to work parallel. For Condition R, however we expected learners to not differentiate between 50% and 17%. Instead we predicted them to prefer the category associate: Learning nouns via verbal restrictions potentially motivates learners to be sensitive to semantic categories and to consider category associates as the best alternative to the high-frequency referent.

### Data Analysis, Results, & Discussion

Learning rates (= high-frequency candidate chosen in Test Type 1) were significantly above chance (25%) for both conditions (N:  $t(23) = 7.995, p < .001$ ; R:  $t(23) = 16.284, p < .001$ ).

More crucially, there were differences in the chosen meaning between conditions in both test types. For analyzing the binomial data of test decisions (low-frequency object chosen vs. high-frequency object chosen), we conducted logistic regressions by entering this binomial data into linear mixed effect models with logit link function (from the lme4 package in R, Bates, 2005). Participant and Item were considered as random factors. To see whether factor Condition had a main effect on test decisions, we compared between the models that include and exclude this factor with a Chi-Square test (Baayen, Davidson, & Bates, 2008). Contrasts between levels (Conditions N and R) were investigated by studying the ratio of regression coefficients and standard errors since the p-values produced by lmers (Wald z test) are anti-conservative (Baayen et al., 2008): If the coefficient is greater than the standard error times two, the comparison is considered to be reliable. The formulas describing the lmer models are of the following form: dependent variable (MeaningChoseb) is a function of (~) the independent variable (Condition) random effects.

In Test Type 1 (high-frequency candidate and low-frequency candidate present) the high-frequency object was selected significantly more often in both conditions. However, it was chosen reliably more often in Condition R than in Condition N (Table 2, Rows 1-2) and the low-frequency ob-

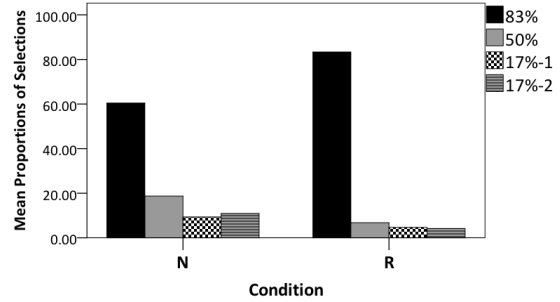


Figure 2: Chosen meaning, Exp. 2, Test Type 1

ject was picked reliably more often in Condition N than Condition R (Table 2, Rows 3-4) (see Figure 2). We also found that confidence ratings were reliably higher in Condition R (7.0) than Condition N (5.5;  $\chi(1) = 31.01, p < .001$ ).

Table 2: Lmer models for chosen meanings in conditions, Test Type 1, Exp. 2  
 $chosen \sim 1 + condition + (1|sub) + (1|item), family = binomial(link = "logit")$

	Predictor	Coef.	SE	Wald z	p
<i>high-freq. object choices</i>					
1	(Int) (N)	0.465	0.188	2.474	< .050
2	R	1.312	0.254	5.160	< .001
<i>low-freq. object choices</i>					
3	(Int) (N)	-1.460	0.185	-7.891	< .001
4	R	-1.151	0.342	-3.369	< .001

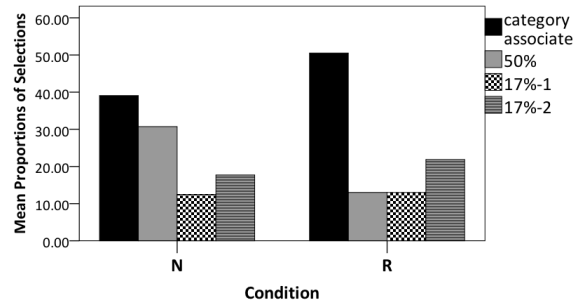


Figure 3: Chosen meaning, Exp. 2, Test Type 2

For Test Type 2 (low-frequency object and category associate available), we also found a remarkable pattern: We found significantly more category-associate (CA) decisions in Condition R than Condition N (Table 3, Rows 1-2) and reliably more low-frequency choices in Condition N than Condition R (Table 3, Rows 3-4) (see Figure 3). To compare whether both the category associate and the low-frequency object were selected significantly more often than each other object, we further conducted repeated measures ANOVAS with Chosen Meaning (CA, low-frequency object, Distrac-

tor 1, Distractor 2) as independent variable. We found main effects for both conditions (N:  $F_1(3, 69) = 9.938, p < .001$ ;  $F_2(3, 45) = 9.018, p < .001$ ; R:  $F_1(3, 69) = 15.165, p < .001$ ;  $F_2(3, 45) = 22.132, p < .001$ ). Pairwise comparisons reveal that, in Condition R, the category associate was selected significantly more often than the three other objects (Table 4, Rows 7-9) and the low-frequency candidate was not chosen more frequently than the distractors (Table 4, Rows 10-12). In Condition N, in contrast, both the category associate and the low-frequency object were selected significantly more often than the two distractors (Table 4, Rows 2-3 and 5-6) but the difference between them was not significant (Table 4, Rows 1 and 4).

Table 3: Lmer models for chosen meanings in conditions, Test Type 2, Exp. 2  
 $chosen \sim 1 + condition + (1|sub) + (1|item), family = binomial(link = "logit")$

	Predictor	Coef.	SE	Wald z	p
<i>category-associate choices</i>					
1	(Int) (N)	-0.543	0.264	-2.058	< .050
2	R	0.605	0.223	2.711	< .010
<i>low-freq. object choices</i>					
3	(Int) (N)	-0.830	0.169	-4.923	< .001
4	R	-1.084	0.267	-4.050	< .001

Table 4: Pairwise comparisons for ANOVAs by subject (Bonferroni adjustment) between category associate (CA) & low-frequency (50%) object vs. each other and distractors (17% objects), Test Type 2, Exp. 2

	chosen	chosen	Mean Diff.	SE	p
<i>Condition N</i>					
1	CA	50%	.083	.074	= 1.00
2	CA	17%-1	.266	.058	< .010
3	CA	17%-2	.214	.060	< .050
4	50%	CA	-.083	.074	= 1.00
5	50%	17%-1	.182	.045	< .010
6	50%	17%-2	0.130	.042	< .050
<i>Condition R</i>					
7	CA	50%	.375	.079	< .010
8	CA	17%-1	.375	.079	< .010
9	CA	17%-2	.286	.089	< .050
10	50%	83%	-.375	.079	< .010
11	50%	17%-1	.000	.032	= 1.00
12	50%	17%-2	-.089	.046	= .402

To summarize Experiment 2, we firstly found a clear sensitivity for differences in the co-occurrence rate of objects and nouns (83% vs. 50% and 50% vs. 17%) in Condition N, which, in contrast, was completely blocked in Condition R. This suggests, firstly, that CSWL works in a parallel manner when it is the only mechanism used but, sec-

ondly, that sentence-level constraints reduced this sensitivity. We attribute the blocking effect to the deterministic nature of the verb cue: Since verb constraints offer a more direct cue, learners relied on its information, ignoring fine-grained co-occurrence relations. Moreover, decisions in trials of Test Type 2 reveal that while learners were more likely to select the category associate than the distractors in both conditions (probably due to the obviousness of the two categories), the difference between the number of category-associate choices and the number of 50%-object choices was only significant in Condition R. This suggests that sensitivity for category associations was enhanced by SLCL. Finally, the difference in confidence ratings in Test Type 1 between conditions reveals that learners were more confident when sentence-level constraints were available than when only statistical information could be used.

## Summary & General Discussion

Results from the two language-learning experiments presented in this paper shed light on the complex interplay of two word-learning mechanisms: Cross-situational word learning (CSWL) and learning based on sentence-level constraints SLCL. Our findings reveal, firstly, that when SLCL and CSWL are in conflict, they have a similar impact on word learning (Experiment 1). Secondly, we found that CSWL-learners are sensitive to small differences in co-occurrence frequencies; however, SLCL clearly blocks this sensitivity when CSWL and SLCL are independently available (Experiment 2). These results about the way CSWL and SLCL interact further allow us to draw conclusions about the underlying nature of both mechanisms: While CSWL offers incremental, probabilistic, and parallel learning, SLCL works in a more deterministic manner. Finally, Experiment 2 provides initial evidence that SLCL leads learners to associate semantic categories with novel nouns, more so than CSWL. This suggests that the two mechanisms result in qualitatively different representations of an emerging word meaning: CSWL yields a set of probabilistically weighted word-meaning mappings, while SLCL associates (presumably verb-derived) semantic features with novel words.

## Acknowledgments

The research reported of in this paper was supported by IRTG 715 "Language Technology & Cognitive Systems" funded by the German Research Foundation (DFG).

## References

- Alishahi, A., & Fazly, A. (2010). Integrating syntactic knowledge into a model of cross-situational word learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 2452–2458). Austin, TX: Cognitive Science Society.
- Altman, G., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73, 247–264.

- Arunachalam, S., & Waxman, S. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*, 442–446.
- Baayen, R., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*, 27–30.
- Field, J. (2004). An insight into listeners' problems: too much bottom-up or too much top-down? *System*, *32*, 363–377.
- Frank, M., Goodman, N., & Tenenbaum, J. (2009). Using speaker's referential intentions to model early cross-situational word learning. *Psychological Science*, *20*, 578–585.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*, 135–176.
- Koehne, J., & Crocker, M. (2010). Sentence processing mechanisms influence cross-situational word learning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 2458–2464). Austin, TX: Cognitive Science Society.
- Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.
- Lee, J., & Naigles, L. (2008). Mandarin learners use syntactic bootstrapping in verb acquisition. *Cognition*, *106*, 1028–1037.
- Quine, W. (1960). *Word and object*. Cambridge, MA.
- Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*, 729–742.
- Yu, C., & Ballard, D. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165.
- Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*, 414–420.
- Yurovsky, D., Fricker, D., Yu, C., & Smith, L. (2010). The active role of partial knowledge in cross-situational word knowledge. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 2609–2615). Austin, TX: Cognitive Science Society.