

Running head: SOCIAL INFORMATION AND CROSS-SITUATIONAL LEARNING

Social information, referential uncertainty, and multiple alternatives tracking in
cross-situational word-learning

(Submitted for the First-Year Project)

Kyle MacDonald

Department of Psychology, Stanford University

Abstract

Word learning proceeds despite the potential for referential uncertainty: within a naming context, a new word could refer to many possible objects, properties of those objects, or even nothing at all. Both social cues and word-object co-occurrence statistics can reduce this uncertainty. But how do learners integrate these two information sources? In a large-scale experiment, we show that the presence of social cues during labeling modulates learners' tracking of cross-situational statistics: without social cues, learners are more likely to track multiple word-object links. Further, I present a computational model that takes a first-step in formalizing how social information influences cross-situational word learning. Together, the data and model suggest that there is a rational tradeoff between reducing uncertainty and the type of representations that learners store.

Introduction

Language is a powerful tool that allows us to rapidly learn from others and building a large early vocabulary creates the foundation for later language and academic success (Hart & Risley, 1995; Qian, 2002). But, to learn the meaning of a new word ¹, children must solve the core problem of referential uncertainty (Quine, 1960): that the range of potential referents for a new word is largely unconstrained. So how do learners reduce this uncertainty?

Social-pragmatic theories argue that the social context of word learning simplifies the word-object mapping problem (Bloom, 2002; Tomasello & Tomasello, 2009). Associative approaches suggest that the rich structure of language allows children to learn new words by tracking word-object co-occurrences over time (i.e., cross-situational learning) (Smith & Yu, 2008). Importantly, these two information sources function at different timescales, with social cues reducing uncertainty during an individual naming event and cross-situational statistics reducing uncertainty across multiple naming events. Moreover, within the cross-situational learning literature, there is considerable debate about the underlying mechanisms that support this type of learning: do learners track information about multiple referents or focus on a single candidate referent?

In this paper, we explore how learners integrate social and cross-situational information and how this interaction modulates the type of representation that learners store. First, I will review evidence in support of children’s ability to use both social cues and cross-situational statistics to reduce uncertainty during word learning. Then I will present the debate over the type of representations that learners store during cross-situational word-learning. Finally, I will present data from a large-scale experiment and a computational model that together show a tradeoff between the presence of social

¹Here I focus on the task of mapping concrete nouns to objects, assuming that the child has solved other requisite inferential problems, e.g., word segmentation.

information and the type of cross-situational information that learners store.

Reducing referential uncertainty

Social cues. Social-Pragmatic theories argue that children's ability to understand a speaker's communicative intent is foundational to word-learning (Bloom, 2002; Baldwin, 1995; Tomasello & Tomasello, 2009). Once the child is capable of rapidly inferring a speaker's communicative intent, the word to object mapping problem is simplified. Experimental work shows the sophistication of children's early intention-reading skills (Baldwin, 1993; Baldwin & Moses, 2001). One powerful demonstration is a series of studies conducted by Tomasello and colleagues in which children had to track and integrate multiple cues to reference in order to successfully learn new words (Akhtar, Carpenter, & Tomasello, 1996; Tomasello, Strosberg, & Akhtar, 1996; Tomasello & Barton, 1994). For example, in Tomasello et al. (1996), children played two versions of a "finding game." In the first version, an adult announced her intention to "find the toma" and then searched in a row of buckets, all containing novel objects. On some trials she would find the object in the first bucket, smile, and terminate her search. On other trials, she would reject the first object by scowling, replacing it in the box, and continuing her search. Critically, 18- and 24-month olds mapped "toma" to the object that the adult *intended* to find, regardless of how many distractor objects were rejected during the search.

Single vs. multiple referent tracking

In the second version, children were familiarized with four objects hidden at four different locations, one of which was a particularly salient toy barn. Once the child had linked each object to its hiding place, an adult announced her intention to "find the gazzer." She then went to the hiding location of the toy barn, but it was locked and she frowned upon not being able to open it. She then continued her search to one of the other

hiding locations, saying "Let's see what else we can find," and pulled out a distractor object while smiling. Interestingly, 18- and 24-month old children mapped the novel word "gazzer" to the toy barn, even though they had not seen the barn after hearing the word and the adult had smiled at a distractor object.

The common thread across all of these studies is that children could not have relied on saliency and co-occurrence information to learn the novel word meanings because the words and objects did not occur close in time and the distractor objects were often more salient. In addition, children were able to use social cues flexibly to determine referential intent: In the first finding game, a smile indicated referential intent, but in the second game, a frown was the relevant cue. Thus, even at 18 months, children are able to infer adults' referential intent and use this information to guide their word-learning in non-ostensive contexts.

Joint Attention. Social Pragmatic theories also emphasize the importance of episodes of joint attention for early word-learning. These episodes involve a caregiver and child attending to an object and understanding that the other participant is also attending to that same object (Carpenter & Tomasello, 1995; Tomasello & Todd, 1983). These contexts are highly structured, repetitive, and limited to a handful of referential intentions, thus simplifying the child's inference. Experimental data show that joint attention facilitates language acquisition (Brooks & Meltzoff, 2008; Farrar, Friend, Forbes, et al., 1993). More recent research suggests that joint attention episodes, in addition to making referential intent clear, provide the child with clear visual access to objects during naming, creating an ideal word-learning situation (Yu & Smith, 2012). Correlational data also show strong links between early joint attention skills (e.g. pointing and responding to parental bids for attention) and later vocabulary growth in typically-developing (Tomasello & Farrar, 1986; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998; Farrar & Zubrick, 2012) and children diagnosed with autism (Mundy, Sigman, & Kasari, 1990).

Additional evidence in support of the importance of rich social interactions for language acquisition comes from recent research investigating the relationship between children's early language environments and language skills. Weisleder and Fernald (2013) measured children's language processing skills, vocabulary, and recorded parent-infant interactions at home using a digital audio recorder that could capture up to 16 hours of speech. These audio recordings were coded for the amount of speech directed to the child and the amount of speech overhead by the child. Importantly, infants who experienced more child-directed speech, but *not overhead speech*, were more efficient in processing language and had larger vocabularies.

Together, there is sufficient evidence that rich, social interactions are high in information value and facilitate early word-learning. However, these episodes make up only a portion of the word-learning contexts that young children experience (Frank, Tenenbaum, & Fernald, 2013) and children can learn language without experiencing much direct social interaction with adults (Shneidman & Goldin-Meadow, 2012). So how might children, in the absence of rich, social information, learn the meanings of words?

Associative word-learning

Associative accounts of word-learning highlight children's powerful statistical learning mechanisms, which track the regularity present in natural language, allowing the learner to map words to objects over time. Researchers in support of this proposal point out that children in early stages of word-learning, before 18 months of age, often have difficulty disambiguating referential intent and infants at much younger ages (8 months) show sensitivity to statistical information in both the auditory (Saffran, Aslin, & Newport, 1996) and visual (Fiser & Aslin, 2002) domains. In this account, word-learning is best explained by domain-general pattern-finding abilities, attention, and memory and words are often mapped to the most salient object in the visual scene.

Empirical data show that in the absence of social cues to word meaning, adults and children are able to rapidly learn words by tracking the co-occurrence of labels and objects across multiple, ambiguous exposures (Smith & Yu, 2008; Vouloumanos, 2008). For example, Smith and Yu (2008) taught children three novel words simply by repeating consistent novel word-object pairings across 10 ambiguous exposure trials. Each exposure consisted of two novel objects and one novel word, with no social cues to reference. Importantly, there was no way for children to infer the correct meaning of the novel word on any one trial, but over time, they accumulated statistical information that led them to make the correct mapping. These results provide strong evidence of children's ability to track the co-occurrence statistics of words and objects in the service of word-learning.

In summary, a substantive body of research supports both the Social-Pragmatic and Associative accounts of early word-learning. However, word-learning requires solving both social-cognitive and mapping challenges simultaneously. Thus, it is likely that children use multiple skills and a variety of information sources to learn words, both social and data-driven. In the next section, I will review experimental and computational work that attempts to take this complexity into account and integrate Social-Pragmatic and Associative accounts.

Integrating Social and Associative Learning

Integrated accounts of early word-learning emphasize the impact of multiple factors on children's ability to learn new words. For example, Hollich et al. (2000) propose an Emergentist Coalitional Model in which they suggest that children take advantage of a "coalition" of cues - social, associative, and linguistic - to learn words. Importantly, these cues are always available to the learner and what changes is how children use each cue. Empirical work provides support for the idea that children do use different sources of information to learn words and the use of specific types of cues changes with development.

Using a preferential looking task, Hollich et al. (2000) asked children aged 10, 12, 19, and 24 months to learn novel words for objects, and pitted social and perceptual cues against each other to test the usefulness of each information source for making the correct object-word mapping at each age. Interestingly, the youngest children ignored the social cue of experimenter eye gaze (i.e., child-guided joint attention), preferring to map the novel word to the more visually salient object (in line with predictions made by the Associative account); however, the older children (20 months and older) preferred to map novel words to the object that the adult labeled.

Recent computational models attempt to formalize how social and statistical cues can be integrated to facilitate word-learning (Yu & Ballard, 2007; Frank, Goodman, & Tenenbaum, 2009). These models, however, treat social information in fundamentally different ways. In Yu and Ballard (2007) Unified Model, social information, such as eye gaze or prosody, is characterized as a "spotlight" that directs the child's attention to objects in the visual scene. In this account, social cues are considered, but intention reading is secondary to attention and saliency. Put another way, social cue simply adds weights to the words and objects that allow statistical learning to operate more effectively.

In contrast, (Frank et al., 2009) propose an *Intentional* model that characterizes social information as providing evidence of a speaker's referential intent, which in turn facilitates word-learning. This model assumes that the relationship between words and objects is mediated by the speaker's goals to refer to some object. By making inferences about a speaker's referential intent, the model is able to simultaneously solve both the social-cognitive and mapping inference problems inherent to word-learning. In addition, Frank and colleagues were able to capture many seminal findings in the early word-learning literature, including cross-situational learning and the use of inferred intentions to disambiguate reference.

In the final section, I propose a new study that follows this line of empirical work

and modeling (Johnson, Demuth, & Frank, 2012; Frank et al., 2009; Yu & Ballard, 2007; Hollich et al., 2000) and explores the interaction between social and statistical learning.

Current Work

Several studies show that in the absence of social cues to word meaning, adults and children are able to rapidly learn words by tracking the co-occurrence of labels and objects across exposures (i.e. cross-situational learning) (Smith & Yu, 2008; Vouloumanos, 2008). However, some researchers question the psychological plausibility of these gradualist accounts, suggesting that children’s rapid word learning is better described by a single hypothesis tracking mechanism (Trueswell, Medina, Hafri, & Gleitman, 2013; Medina, Snedeker, Trueswell, & Gleitman, 2011).

Recent experimental work shows that both adults and children can track and recall multiple referents (Yurovsky & Frank, in prep). In Yurovsky and Frank’s task, participants saw a set of novel objects and heard a novel word (e.g. Grink), and were asked to make a guess about the “correct”² word-object mapping. In subsequent test trials, participants heard the novel word again, this time paired with another set of novel objects. Critically, one of the objects in the set was either the participant’s initial hypothesis (Same trials) or one of the objects that was *not* the initial hypothesis (Switch trials). On Switch trials, adults reliably selected the object that was not their initial hypothesis, even when there were eight objects in the initial exposure set, providing strong evidence that learners track multiple referents when learning new words.

However, this task did not include any of the rich, social cues that typically accompany real world word-learning. So it is still an open question as to how social information interacts with learners’ demonstrated ability to track multiple referents.

²There was actually no “correct” answer on exposure trials; rather these trials gave participants evidence about potential word-object mappings and allowed them to form an initial hypothesis.

Perhaps the presence of additional evidence about a speaker’s referential intent strengthens the learner’s initial hypothesis, reducing the need to track alternative hypotheses. Or social information could strengthen the initial hypothesis without reducing multiple referent tracking. The proposed study asks if the presence of social information changes how learners track multiple referents when learning new words?

The answer to this question will increase our understanding of how social information interacts with statistical learning mechanisms with the overarching goal of clarifying the unique contribution of each information source to children’s early word learning.

Experiment

Participants

This experiment was posted to Amazon Mechanical Turk as a set of Human Intelligence Tasks (HITs) to be completed only by participants with US IP addresses that paid 30 cents each (for a detailed comparison of laboratory and Mechanical Turk studies see Crump, McDonnell, & Gureckis, 2013). 30 HITs were posted for each condition (Social, Non-social) for total of 60 paid HITs. If a participant completed the experiment more than once, he or she was paid each time but only data from the first HITs completion was included in the final data set. In addition, data was excluded from the final sample if participants did not give correct answers for familiar trials (excluded 1 HIT).

Stimuli

Stimuli for the experiment consisted of black and white pictures of familiar and novel objects, a schematic drawing of a human interlocutor, and audio recordings of familiar and novel words. Pictures of 32 familiar objects spanning a range of categories (e.g. squirrel, truck, tomato, sweater) were drawn from the set constructed by Snodgrass

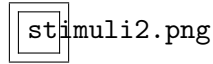


Figure 1. A schematic of the trials that participants saw in the experiment. On Exposure trials, participants saw four novel objects and heard a novel word. Participants were asked to guess it's correct referent. After the Exposure trial, participants either saw a Same or a Switch trial. On Same trials, the set of four referents contained the participant's previous hypothesis. On Switch trials, the set of referents contained one of the objects that the participants had *not* hypothesized.

and Vanderwart (1980). Pictures of distinct but difficult to name objects were drawn from the set of 140 first used in Kanwisher, Woods, Iacoboni, and Mazziotta (1997). For ease of viewing on participants' monitors, pixel values for all pictures were inverted so that they appeared as white outlines on black backgrounds (see Figure 1). Familiar words consisted of the labels for the familiar objects as produced by ATT Natural Voices™ (voice: Crystal). Novel words were 1-3 syllable pseudowords obeying the rules of English phonotactics produced using the same speech synthesizer.

A schematic drawing of a human interlocutor was chosen for ease of manipulating the direction of eye gaze, the social cue of interest in this study. Five images were created using Adobe Photoshop with the following directions of eye gaze: far left, close left, close right, far right, and eyes center. Because this task was performed over the internet and participants' screen sizes might be small, it was important to make the eye gaze cue clear.

Design and Procedure

Participants were exposed to a series of trials in which they heard an interlocutor say a novel word, saw four novel objects, and were asked to indicate their guess as to which object was the referent of the word. After a written explanation of this procedure,

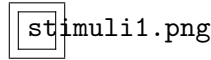


Figure 2. Exposure trials in the Social condition. In the Nonsocial condition, the interlocutor looked straight ahead during exposure. During test trials, the interlocutor looked straight ahead.

participants were given four practice trials to introduce them to the task. On each of these trials, they heard the interlocutor say a Familiar word while looking at a line drawing of that object among a set of other familiar objects. On the first two trials, participants were asked to find the squirrel, and the correct answer was in the same position on each trial. On the next two trials, participants were asked to find the tomato, and the correct answer switched positions from the first to the second trial (in order to ensure that participants understood that the on-screen position was not an informative cue to the correct target). These trials also served to screen for participants who did not have their audio enabled or who were not attending to the task.

After these Familiarization trials, participants were informed that they would now hear novel words, and see novel objects, and that they should continue selecting the correct referent for each word. Participants each heard eight novel words twice. Participants saw four referents on each trial, and the two trials for each word occurred back-to-back. Four of these follow-up trials were Same trials in which the referent that participants selected on the exposure trial appeared again amongst the set of objects. The other four were Switch trials in which one of the referents in the set was selected randomly from the objects a participant did not select on the previous Exposure trial. All other referents were completely novel on each trial. Critically, on Exposure trials the interlocutor's eye gaze was directed and thus informative, but on Same/Switch trials her eye gaze was undirected and thus uninformative.

Participants were randomly assigned to one of two conditions: Social or Nonsocial. In the Social condition, eye gaze was informative on exposure trials. In the Nonsocial condition, eye gaze was uninformative throughout the task, i.e. the speaker always looked straight ahead.

Because participants performed this task over the internet, it was important to indicate to them that their click had been registered. Thus, a red dashed box appeared around the object they selected for 1 second after their click was received. This box appeared around the selected object whether or not it was the "correct" referent.

Model

First, I begin by describing the computational model developed by Frank et al. (2009) and extended by Yurovsky and Frank (in prep). Then, I will discuss how I modified the model's parameters in order to generate predictions about the effect of social cues to reference on cross-situational learning.

In Frank et al.'s model, the word-learning problem is defined by a set of variables and probabilistic dependencies that link these variables. The learner observes a set of situations S , which consist of two observed variables - objects (O) and words (W) - and one hidden variable: the speaker's intention (I), with the goal of inferring the most probable lexicon of word-object mappings (L) given the set of observed situations. Formally, the task of learning the correct lexicon can be defined as a problem of Bayesian inference:

$$P(L|S) \propto P(S|L)P(L) \quad (1)$$

The likelihood term ($P(S|L)$) captures the learner's assumptions about the task. Each situation (S) includes a word, a set of objects, and the speaker's referential intention (I).

Thus, $P(S|L)$ can be expanded and the probability of the lexicon can be given as:

$$P(L|S) \propto \prod_{s \in S} P(W_s, I_s, O_s|L)P(L) \quad (2)$$

Because referential intention mediates the relationship between words and objects, the likelihood term can be rewritten as:

$$P(L|S) \propto \prod_{s \in S} P(W_s|I_s, L)P(I_s|O_s)P(L) \quad (3)$$

The prior probability of a lexicon is defined as: $P(L_o) \propto \frac{1}{|L_o|}$, which is a parsimony prior that favors lexicons with fewer words referring to the same object.

Thus far, I have described the computational problem for the Bayesian word learner outlined by Frank et al., 2009. Yurovsky and Frank then extended this computational level description to include cognitive constraints (memory and attention), allowing them to provide an algorithmic-level account of their data. Memory was modeled as a power function that took into account the interval between successive exposures to a word. The γ parameter scaled to the strength of the initial encoding and the λ parameter defined the rate of decay.

$$M(L_o) = \gamma L_o t^{-\lambda} \quad (4)$$

Attention was defined in terms of the learner’s belief about $P(I|O)$: the probability of the speaker’s intention to refer to each object in the set. The σ parameter captures the amount of probability mass that the learner places on each referent during the initial exposure. The strength of the learner’s belief about $P(I|O)$ in turn directly influences the strength of initial encoding. At $\sigma = 1$, the learner is placing all of the probability mass on one word-object mapping (i.e. a single hypothesis tracking strategy). At $\sigma = \frac{1}{|O|}$, the learner is distributing the probability mass across all possible word-object mappings equally (i.e. a parallel statistical accumulator strategy).

Figure 1 shows model predictions for four different levels of belief about referential intent (i.e. four different values for σ). The first two models are from Yurovsky and Frank

(in prep). The Parallel Accumulator model, $\sigma = (\frac{1}{|\sigma|})$, predicts indistinguishable performance on Same and Switch trials because the learner is allocating belief evenly across referents. The No Social Integrated model predicts strong performance on Same trials and above chance performance on Switch trials because it allocates a chunk of probability to a single hypothesis and distributes the rest to the other referents. This model was the best fit for human behavior on their cross-situational learning task, explaining 98% of the variance in the data.

However, these two models did not include the rich social cues to reference that are available to the word learner. Perhaps more evidence about referential intent will cause the learner to place more belief on the object that is the target of the referential act. To formalize this intuition, I modified the σ parameter in Frank and Yurovsky’s model to generate predictions about human performance on a social, cross-situational learning task (see Figure 1). The Weak Social model treats social information as additional evidence about reference, and increases the allocation of belief to a single hypothesis, $\sigma = 0.75$. It predicts an improvement on Same trials and a decrease in performance on Switch trials. However, the learner still tracks multiple referents and does not place *all* belief on a single hypothesis. In contrast, the Strong Social model ($\sigma = 1$) predicts that learners will not track multiple referents because all of their probability mass is allocated entirely to the target of the speaker’s referential intent, similar to the single-hypothesis tracking strategy suggested by Trueswell and colleagues.

References

- Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, 67(2), 635–645.
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of child language*, 20(02), 395–418.
- Baldwin, D. A. (1995). Understanding the link between joint attention and language.
- Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development*, 10(3), 309–329.
- Bloom, P. (2002). *How children learn the meaning of words*. The MIT Press.
- Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, 35(01), 207–220.
- Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, i–174.
- Carpenter, M., & Tomasello, M. (1995). Joint attention and imitative learning in children, chimpanzees, and enculturated chimpanzees*. *Social Development*, 4(3), 217–237.
- Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PloS one*, 8(3), e57410.
- Farrant, B. M., & Zubrick, S. R. (2012). Early vocabulary development: The importance of joint attention and parent-child book reading. *First Language*, 32(3), 343–364.
- Farrar, M. J., Friend, M. J., Forbes, J. N., et al. (1993). Event knowledge and early language acquisition. *Journal of child language*, 20, 591–591.
- Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99(24), 15822–15826.

- Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5), 578–585.
- Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, 9(1), 1–24.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children*. Paul H Brookes Publishing.
- Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the society for research in child development*, i–135.
- Johnson, M., Demuth, K., & Frank, M. (2012). Exploiting social information in grounded language learning via grammatical reductions. In *Proceedings of the 50th annual meeting of the association for computational linguistics: Long papers-volume 1* (pp. 883–891).
- Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, 9(1), 133–142.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019.
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and developmental Disorders*, 20(1), 115–128.
- Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and

- academic reading performance: An assessment perspective. *Language learning*, 52(3), 513–536.
- Quine, W. V. (1960). 0. word and object. 111e MIT Press.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.
- Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a mayan village: how important is directed speech? *Developmental science*, 15(5), 659–673.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106(3), 1558–1568.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of experimental psychology: Human learning and memory*, 6(2), 174.
- Tomasello, M., & Barton, M. E. (1994). Learning words in nonostensive contexts. *Developmental psychology*, 30(5), 639.
- Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child development*, 1454–1463.
- Tomasello, M., Strosberg, R., & Akhtar, N. (1996). Eighteen-month-old children learn words in non-ostensive contexts. *Journal of child language*, 23, 157–176.
- Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, 4(12), 197–211.
- Tomasello, M., & Tomasello, M. (2009). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive psychology*, 66(1), 126–156.

- Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, 107(2), 729–742.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological science*, 24(11), 2143–2152.
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70(13), 2149–2165.
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*.