

Running head: SOCIAL INFORMATION AND CROSS-SITUATIONAL LEARNING

The role of social information in cross-situational word-learning

(Submitted for the First-Year Project Proposal)

Kyle MacDonald

Department of Psychology, Stanford University

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). However, to successfully learn a new word, children must solve several problems simultaneously. They must be able to: interpret a speaker’s utterance as having communicative intent, extract individual words from that utterance, map that word to the correct object in the world, and retain that link over time. Together, these social-cognitive and mapping challenges make word-learning a surprisingly difficult puzzle for the child to solve. So how do children learn words so rapidly?

Different theories of early word-learning emphasize different learning tools and information sources available to the child. These proposals can be divided into two broad categories: *Social-Pragmatic* and *Associative* accounts. Social-Pragmatic theories characterize language acquisition as a fundamentally social process, with word-learning occurring as a result of the interaction between children’s social-cognitive skills and highly structured learning moments (often labeled episodes of *joint attention*). In contrast, Associative proposals highlight children’s powerful statistical learning mechanisms and the rich structure of natural language. In this account, word-learning is best explained by domain-general pattern-finding abilities, attention, and memory.

My goal in this paper is to briefly review the evidence in support of these two proposals (Section 2) as well as recent experimental and computational work that attempts to integrate the two (Section 3). In the final section (Section 4), I will propose a new study that explores how social information interacts with associative learning mechanisms during cross-situational word-learning.

Social-Pragmatic word-learning

Intention-reading Skills. 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.

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 tries to explore the interaction of social and statistical learning.

Proposed Research

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”¹ 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. 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

¹There 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.

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.

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.
- 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).
- 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.
- 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.
- 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*.