Continuity of Discourse Provides Information for Word Learning

Michael C. Frank, Noah D. Goodman, and Joshua B. Tenenbaum
{mcfrank, ndg, jbt}@mit.edu
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology

Anne Fernald
afernald@stanford.edu
Department of Psychology
Stanford University

Abstract

Utterances that are close in time are more likely to share the same referent. A word learner who is using information about the speaker's intended referents should be able to take advantage of this continuity and learn words more efficiently by aggregating information across multiple utterances. In the current study we use corpus data to explore the continuity of reference in caregivers' speech to infants. We measure the degree of referential continuity in two corpora and then use regression modeling to test whether reference continuity is informative about speakers' referential intentions. We conclude by developing a simple discourse-continuity prior within a Bayesian model of word learning. Our results suggest that discourse continuity may be a valuable information source in early word learning.

Keywords: Language acquisition; discourse; word learning; Bayesian modeling

Introduction

Imagine attending a dinner party where you don't speak the language very well. Most of the time you will likely have trouble understanding what the conversation is about, and if you don't understand what is being talked about, you will have a hard time guessing the meanings of new words. There may be opportunities, however, where you can guess the topic of conversation and infer some word meanings. For example, if a guest gestures towards her dinner plate as she makes a comment, you can guess that the topic of conversation is the food and perhaps that one of the words she used means "trout."

The problem of word learning has a similar structure for children learning their first language. If they are engaged in a joint activity or even a moment of joint attention, they can use this information to make inferences about the speakers' referential intentions and the meanings of words (Tomasello, 2001). Our recent computational work has elaborated this idea—that inferring the intentions of a speaker can give a sophisticated word learner leverage in inferring the meanings of the words the speaker uses (Frank et al., in press).

Going back to our dinner party, a learner who assumes the guest's utterance is about the trout is making use of immediate contextual information about the speaker's intentions. But if a second guest speaks up immediately afterwards, the learner could guess with some certainty that this remark also has to do with the trout (or if not, at least the potatoes or the salad). This kind of aggregation of information across time makes use of the continuity of discourse in conversation. If the second guest's remark had come an hour or even a minute after the first remark, the learner would have had much more uncertainty about the topic of conversation.

Discourse structure has been well-studied in psycholinguistics (Graesser et al., 1997). Despite this—and despite the potential utility of discourse information in word learning, as illustrated by the dinner party example—research on word learning has largely neglected the role of discourse continuity. For example, although a number of recent computational models use cross-situational information about the co-occurrence of words and referents for word learning, most of these models assume that utterances are sampled independently from one another with respect to time, throwing away important information about the order of utterances (Siskind, 1996; Yu & Ballard, 2007; Frank et al., in press).[1]

Our goal in the current paper is to investigate the utility of discourse continuity for word learning. Although a more elaborate model of discourse would contain abstract topics like "the quality of the food served in the main course," here we consider a very simplified version of talking about the same topic: talking about the same object (continuity of reference), which is common in interactions with infants. Although this approach may throw away more abstract information about the kind of activity or action that the child and caregiver are jointly involved in, it is more likely to be the kind of information available to even the youngest word learner. For the rest of the paper, we will use the terms "continuity of reference" and "discourse" interchangeably.

The plan of the paper is as follows. We first introduce the corpora we studied. We next discuss a simple model of how to measure about continuity of reference within these corpora and use a supervised (regression) model to test whether discourse continuity provides information about speakers' referential intentions. We then end by creating a simple prior distribution over referential intentions which favors continuity of reference and applying it within our model of intentional word learning.

[1] But cf. Roy & Pentland (2002), who used a recurrence filter to take into account temporal information.

Figure 1: A sample frame from the FM corpus.

Corpora and coding

For our initial analysis of discourse continuity, we chose to study corpora of child-caregiver interactions by annotating them with information relevant to discourse. For our analysis, we chose corpora based on two criteria. First, corpora needed to include video as well as audio so that we could accurately identify both the speaker's referential intentions (the objects they were talking about) and the objects present in the physical context. Second, corpora needed to be collected in a restricted enough context that it would be feasible to code the entire set of plausible referents for a word, so that the set of alternative referents for a word could be considered.

We selected two corpora which fulfilled these requirements. The first was a pair of two 10 minute videos from the CHILDES Rollins corpus (`me03` and `di06`) (MacWhinney, 2000). These videos recorded mothers interacting with preverbal infants by selecting toys from a larger set. The videos contained 316 and 303 utterances which made reference to 21 and 18 toys, respectively. This corpus was previously used in several computational studies of cross-situational word learning (Yu & Ballard, 2007; Frank et al., in press).

The second corpus was a larger set of videos of object-centered play between mothers and children in their homes, collected by Fernald & Morikawa (1993). We refer to this as the FM corpus. Although the original study considered videos of American and Japanese mothers, in the current study we only made use of the American data. There were 24 total videos, ranging from approximately 10 to 30 minutes and containing from 82 to 554 utterances (mean = 311) which made reference to 44 total objects. The children in these videos fell into three age groups: 6 months (N=8), 11-14 months (N=8), and 18-20 months (N=8). Mothers in the videos were given several pairs of toys by the experimenter and asked to play with each pair for a 3-5 minute period; thus, similar to the Rollins corpus, the total set of objects present in the videos was severely restricted.

We operationalized "referential intention" as an intention to refer linguistically to an object. We coded an utterance as referring to an object when the utterance contained either the name of the object or a pronoun referring to that object. For each of the corpora, we coded the referential intention for all of each caregiver's utterances. For example, in a sentence like "look at the doggie," the referential intention would clearly be to talk about the `dog`. Likewise, in an utterance like "look at his eyes and ears," (where the caregiver was pointing at the dog), the referential intention would also be the `dog`—though the coder would need to make reference to the videotape to determine the pronoun reference. We did not mark the use of property terms like "red," super-/subordinate terms like "animal" or "poodle," or part terms like "eye." Exclamations like "oh" were not judged to be referential, even if they were directed at an object. Objects that were not present were still judged to be part of a referential intention, e.g., "do you like to read books" would be judged to have the intention `book` even if the child could not see a book or a book was not present in the scene at all.

In addition to coding the referential intention of the speaker, for each utterance we also coded the mid-sized objects present in the field of view of the learner at the time of the utterance. A sample frame from the FM corpus is shown in Figure 1. The only object judged to be in the field of view of the child at the time of the utterance most proximate to this frame was the `dog`. The end product of this coding effort was two corpora, one of around 600 utterances and one of almost 8000 utterances, for which each utterance was annotated with the objects present in the field of view of the learner and the referential intention(s) of the speaker.

Predicting reference

The first goal of our study was to investigate and describe factors involved in determining whether objects are referred to in caregivers' speech to children. Towards this goal, we first developed a visualization of reference in child-directed speech; we then attempted to quantify the contributions of physical presence, discourse continuity, and discourse novelty to object reference.

Visualizing continuity of discourse

The first step we took towards understanding the prevalence of discourse continuity in the Rollins and FM corpora was to visualize the results of our coding. We introduce what we call a "Gleitman plot": a visualization of a stretch of discourse based on (1) what objects are present and (2) what objects are being talked about.[2] Example Gleitman plots for one file in each of the corpora are shown in Figure 2.

We can draw two anecdotal conclusions on the basis of these visualizations. First, within the corpora we studied, mothers talk primarily about objects that are present in the field of view of the children. This can be seen by examining the spread of green within each plot. The largest bout of green is in the lower of the two plots, when the mother is playing a hiding game with several of the toys. For a word learner

[2]Named because Gleitman (1990) was concerned with the relationship between what is present in a learner's experience and what is being talked about.

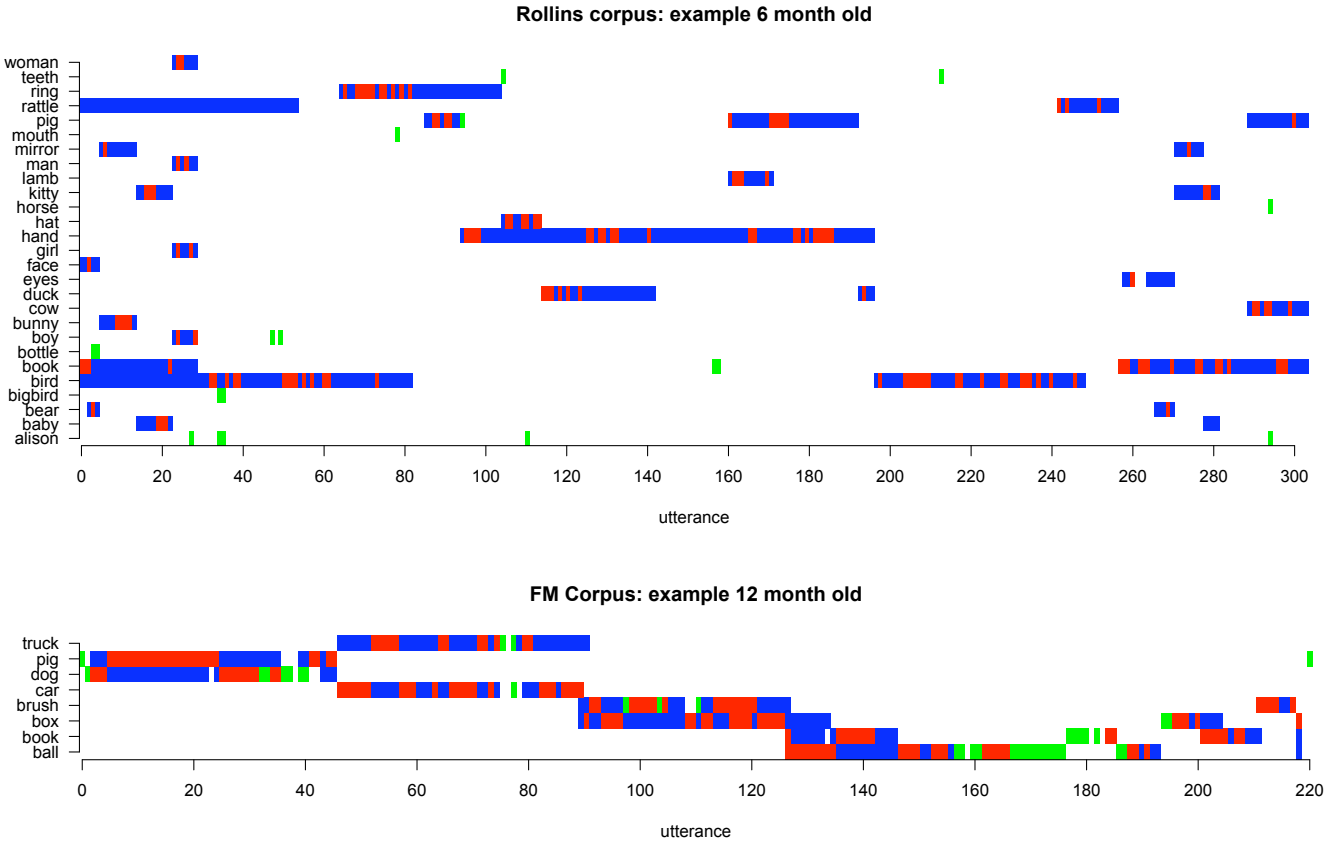

Figure 2: Example Gleitman plots for videos from the Rollins and FM corpora. Each row represents an object, each column represents an utterance. A blue mark denotes that the object was present when the utterance was uttered but not mentioned; a green mark denotes that the object was mentioned but not present; and a red mark denotes that the object was present and mentioned. The streaks of red indicate bouts of continuous utterances referring to a particular object.

guessing the meaning of a novel word, the best guess will likely be that the word refers to an object that is present, although though this generalization may not be nearly as useful when learning verbs rather than nouns (Gleitman, 1990).

Second, we can see clear evidence of discourse continuity in both files. Rather than being distributed evenly throughout the span of time when an object is present, references to an object are "clumpy": they cluster together in bouts of reference to a single object followed by a switch to a different object. This can be seen for example in the `car` / `truck` portion of the FM example (utterances 47 - 67), where the mother alternates several times between the two objects, talking about each object for several sentences before switching.

Quantifying continuity of discourse

In our visualizations of discourse continuity, we observed "clumps" of references to a particular object rather than a more uniform distribution of references over time. To quantify this trend, we first defined a quantitative measure of discourse continuity. For an object o, we defined $R_t(o)$ as a delta function returning whether or not that object was referred to at time t. Next we defined the probability of discourse continuity $P_D(o)$. This measure captures the probability of an object being talked about, given that it was talked about in the sentence before:

$$P_D(o) = \frac{\sum_t R_t(o) R_{t-1}(o)}{\sum_t R_t(o)} \quad (1)$$

We calculated $P_D(o)$ for each object for the times when it was present in the physical context. We then took an average of $P_D(o)$ over all objects, weighted by the frequency of each object, to produce an average value for each file.

We then estimated a baseline value for P_D via permutation analysis. Intuitively, this analysis asks what a "chance" value for P_D would be if utterances were completely independent of one another. We calculated this baseline value for each corpus file by recomputing $P_D(o)$ for 100 random permutations of

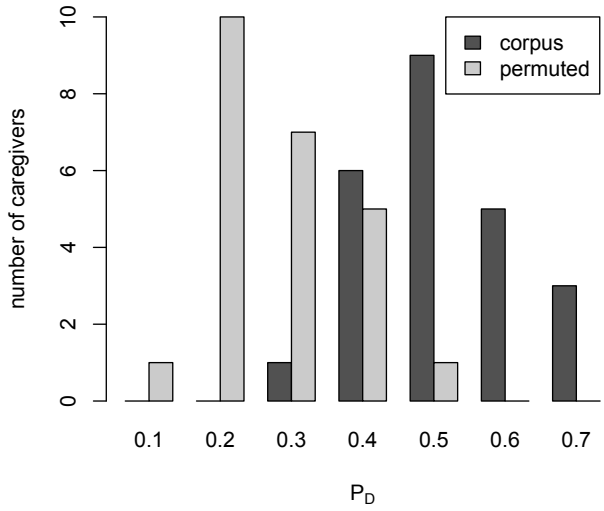

Figure 3: A histogram of the mean value of P_D (probability of discourse continuity) for each file in the FM corpus (dark gray) and for permuted baseline values (light gray).

the times at which each object was talked about.[3] For the Gleitman plots in Figure 2, this analysis would be represented by randomly shuffling all the red and blue squares in each row so that the same overall set of squares were red and blue but their distribution was different.

The results of this analysis for the FM corpus are shown in Figure 3. As predicted based on our visualizations, P_D differed significantly from the permuted baseline (paired $t(23) = 7.85$, $p < .0001$, Cohen's $d = 1.50$). In addition, in a simple linear regression, we found no relationship between P_D and age ($r^2 = 0.067$, $p = .22$). For the Rollins files, the mean values of P_D were .46 and .61 for the `di06` and `me03` files, respectively.

Quantifying discourse novelty

An object's novelty in the context provides an additional factor governing how likely a speaker is to refer to an object. Intuitively, an object that is newly part of the physical context is more likely to be talked about, and some some empirical evidence suggests that children may be able to make use of this information to learn new words. Akhtar et al. (1996) found that two-year-olds were able to use the fact that an object was new to the experimenter (even though the children themselves had already played with it) to infer that the object was the experimenter's intended referent and hence was named by the novel word the experimenter produced.

[3]Excluding utterances during which an object was not present was important in calculating an accurate baseline; had we permuted all utterances, we would have artificially deflated the baseline by spreading references to o across the entire file even when o was not present.

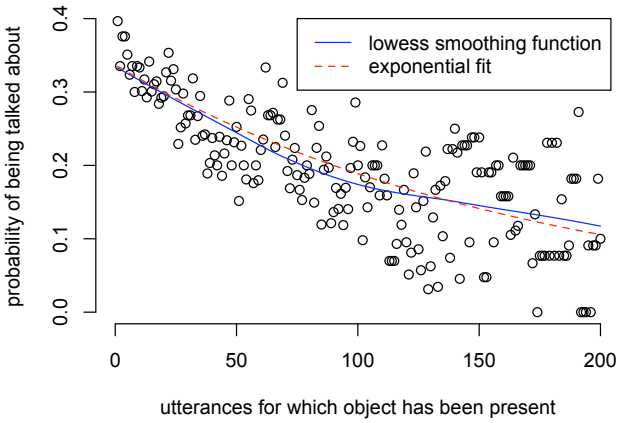

Figure 4: The mean probability of an object being talked about as the number of utterances for which the object was present increases. Data are for the FM corpus.

To quantify the effects of discourse novelty on the probability of talking about an object, we plotted the mean probability that an object was being talked about (given that it was present) by the number of utterances for which the object had been present. We performed this analysis only on the FM corpus, since the Rollins corpus was too sparse to provide accurate estimates. The results are plotted in Figure 4; the resulting curve was well-described by an exponential function, capturing the generalization that the longer an object is present, the less likely it is to be talked about.

Regression modeling

In the previous sections we discussed three factors which contributed to an object being talked about within the corpora we studied: whether it was physically present, whether it was being talked about in the previous sentence, and whether it was relatively new in the context. In our next analysis we set out to quantify their relative contributions to predicting speakers' reference. To do this, we used multi-level logistic regression models (Gelman & Hill, 2006). We selected a logistic regression since the predicted measure—referring to an object or not—was binary, and we used a multi-level model in order to estimate and remove variance due to the effects of different objects and speakers.

We fit a single regression model for each of the two corpora. This model predicted whether an object would be talked about at a particular time and incorporated group-level fixed effects of (1) presence, whether an object was present in the physical context; (2) discourse continuity, whether an object was referred to in the previous utterance; and (3) discourse novelty, whether an object was new in the physical context; along with partially-crossed random effects of caregiver and object. Fixed effects were relatively uncorrelated (pairwise r values less than .34 for all predictors). Random effects served

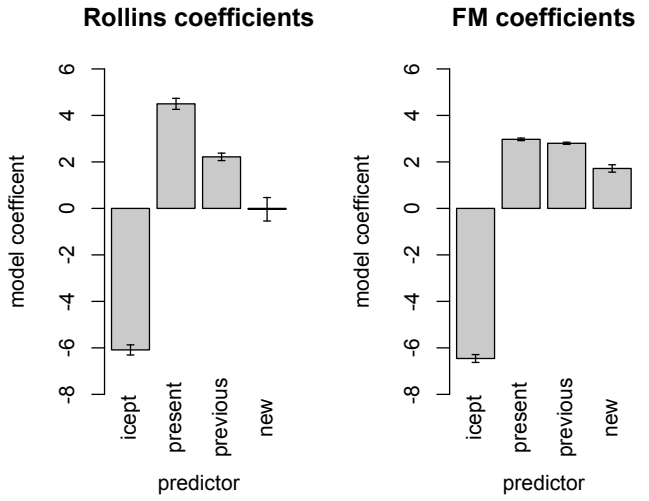

Figure 5: Coefficient estimates for mixed logistic regressions predicting whether an object would be talked about in a particular utterance. Models are for the Rollins corpus (left) and the FM corpus (right). Predictors are (from left to right), the model intercept, whether the object was present, whether the object had been talked about in the previous utterance, and whether this was the first utterance during which the object was physically present.

to remove systematic variation due to differences in how often particular caregivers used referential language and how frequently different objects were referred to.

The results of this analysis are shown in Figure 5. For both corpora, the models had large negative intercepts, indicating a very low likelihood of talking about any given object *a priori*. We saw a highly significant positive coefficient on the object being present in the physical context for both models (both ps $< .0001$), though the difference in coefficient estimates (4.50 for Rollins versus 2.97 for FM) was likely an artifact due to the greater diversity of objects present in the Rollins study. In addition, we saw a highly significant positive coefficient on the discourse continuity term for both models (both ps $< .0001$); these coefficients are comparable and they are very similar in magnitude (2.21 for Rollins and 2.80 for FM). Finally, we saw a difference between the two models in the discourse newness predictor. While this predictor was significant for the FM corpus ($p < .0001$), it did not reach significance for the Rollins corpus ($p = 0.94$), possibly due to data sparsity.

These regression models take a first step towards quantifying some intuitions about the utility of discourse continuity. For the most part, caregivers talk about what is present. If something new has come along, they are likely to talk about it, and if not, they will likely keep talking about what they were talking about a moment ago. While none of these three regularities are hard-and-fast rules, they may allow a learner to make a good guess about what is being talked about in cases which would otherwise be ambiguous.

Using discourse continuity for word learning

The results of the previous analyses suggested that expectations about discourse may help a learner guess what is being talked about. In our final analysis we explored the possibility of using discourse information in an unsupervised model of word learning.

We began with the intentional word learning model described in Frank et al. (in press). This model takes as its inputs a set of situations: utterances and the objects that are present at the time of the utterance. It assumes first that for every situation, the speaker has chosen some subset of the objects in the situation to talk about (possibly an empty subset). These objects comprise her referential intention. The model assumes second that the speaker is likely to utter words that are linked to these objects in the lexicon, in addition to some number of other words that do not refer directly to objects in the situation. These two assumptions and the pattern of co-occurrences between words and objects in the data jointly define a probability distribution over two latent states: the lexicon of the language (the set of mappings between words and objects) and the referential intention of the speaker in each situation.

In the work reported in Frank et al. (in press), solving for the most probable lexicon using Bayesian inference resulted in learning a more accurate lexicon than purely associative models. But in these simulations, situations were assumed to be sampled independently from one another and no other information about referential intentions was given to the model. Therefore, the model assumed a uniform probability distribution over intentions. Our current work suggests that this uniform prior over intentions may be inappropriate. To make a preliminary test of whether altering this assumption might result in more effective learning, we constructed a prior distribution which privileged continuity of intention.

A discourse continuity prior

Our original model assumed that the speaker's referential intention at time t, I_t, was chosen uniformly from all the possible subsets of the objects present at that time ($O_t^\star$, the power set of O_t). To define a prior that takes into account discourse continuity, we create a dependency between I_t and I_{t-1}. We assume that when choosing I_t, the previous intention I_{t-1} can be chosen with probability δ, or a new intention can be chosen uniformly from $O_t^\star$ with probability $1-\delta$. (If the objects in O_{t-1} were not no longer present in O_t, we assumed a uniform choice over the new possible intentions).

By introducing this temporal dependency, the discourse prior converts the original word learning model into a hidden Markov model in which the intention is the hidden state. To score a lexicon during inference, we summed over all possible sequences of intentions using the forward algorithm (Rabiner, 1989).

Simulation Data

We made a preliminary test of our discourse continuity prior by creating a small corpus in which the reference of a novel

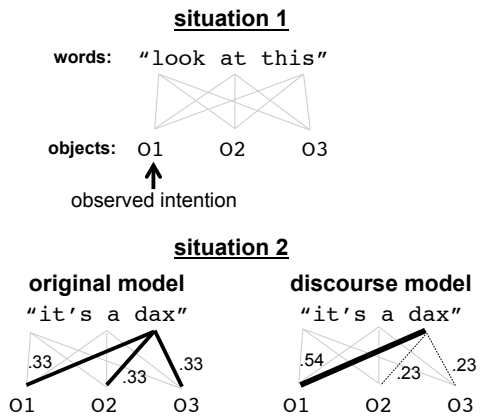

Figure 6: A schematic depiction of the two key situations in our simulation. In situation 2, the relative probabilities of lexicons including different links between the novel words and the objects are shown for the original model and the discourse model.

word was ambiguous, but the speaker's intention was revealed in the previous situation (pictured in Figure 6). One way of conceptualizing this corpus is as a simple, child-directed version of the dinner party example with which we began the paper. A first utterance ("Look at this!"), combined with some sort of clear intentional cue (e.g., a look or a point to the intended object), establishes the discourse referent. Then, the following utterance names the object with a novel word (e.g., "It's a dax."). Other situations, not shown, gave examples of the familiar words (e.g. "look") in a range of other contexts and established that they did not consistently refer to any single object.

We ran both the original model and the model with the discourse prior on this artificial corpus. We found that while the original model was not able to learn the mapping between the novel word and the previously-intended object, the discourse model preferred a lexicon which included this mapping. Thus, including the discourse prior in the model allowed it to make use of the intentional information even though it did not co-occur precisely with the novel word.

General Discussion

We began by suggesting that a word learner could take advantage of the continuity of discourse to aggregate information about speakers' intentions over time and then use better guesses about intention to learn words more effectively. To support this claim we analyzed two corpora of mother-child interactions. We found first, that caregivers' discourse was extremely continuous across a range of ages and situations, and second, that for a supervised learner what a speaker had just talked about was informative about what she was going to talk about. We then added a prior on discourse continuity into our unsupervised model of word learning and found that this prior allowed learning in situations that would otherwise be ambiguous.

Our aim here has been to identify what we believe to be an important source of information for word learning. Work on language understanding has long acknowledged the importance of discourse information (Graesser et al., 1997). In contrast, researchers studying word learning are only beginning to conceptualize this task as language understanding in the presence of uncertainty about the meanings of words. We hope that our work here inspires future research into connections between language understanding and language learning.

Acknowledgments

We gratefully acknowledge the work of Maeve Cullinane in coding the FM corpus. Thanks to Steve Piantadosi and the members of tedlab and cocosci for valuable comments. This work supported by a Jacob Javits Graduate Fellowship to the first author and NSF DDRIG #0746251.

References

Akhtar, N., Carpenter, M., & Tomasello, M. (1996). The role of discourse novelty in early word learning. *Child Development*, *67*, 635-645.

Fernald, A., & Morikawa, H. (1993). Common themes and cultural variations in japanese and american mothers' speech to infants. *Child Development*, *64*, 637–56.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (in press). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55.

Graesser, A., Millis, K., & Zwaan, R. (1997). Discourse comprehension. *Annual Reviews in Psychology*, *48*, 163–189.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*, 257–286.

Roy, D., & Pentland, A. (2002). Learning words from sights and sounds: a computational model. *Cognitive Science*, *26*, 113–146.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–91.

Tomasello, M. (2001). Perceiving intentions and learning words in the second year of life. In *Language acquisition and conceptual development* (pp. 132–159). New York: Cambridge University Press.

Yu, C., & Ballard, D. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165.