

# **Early Word Learning Through Communicative Inference**

## **Dissertation Summary**

**Michael C. Frank**

How do children connect words with their meanings? As Quine observed, for any body of evidence about the use of a word, there are an infinite possible range of meanings that could fit the evidence exactly. Work on early word learning has recognized many ways that learners can overcome the problem of referential indeterminacy, from conceptual heuristics, to explicit social signals, syntactic cues, or cross-situational associations. Yet reviews of the state of the art in word learning read like a list of ad hoc strategies, ending with an acknowledgment that future work needs to bring together disparate proposals into a more coherent framework.

More generally, even though these word learning strategies are well-supported by the empirical data, there is still disagreement about their underlying mechanisms. For example, researchers have argued that early word-object mappings are learned primarily through fast, socially-guided inferences or primarily through the slow growth of associative linkages. Both of these kinds of learning can be elicited in the lab but it is unknown how they are related to each other and to the other kinds of inferences learners make in language acquisition. This thesis presents a computational and theoretical framework for unifying these two influential sets of ideas and sets of empirical results by suggesting that early word learning can best be described as a process of joint inferences about speakers' referential intentions and the meanings of words. Taking this perspective has the promise to explain and unify a wide range of findings.

**Chapter 1** of the thesis presents a wide-scope overview of the role of probabilistic inference in language acquisition. A wealth of recent evidence suggests that probabilistic

learning mechanisms play a role in language acquisition. Nevertheless, the structure of these mechanisms is controversial and it is still unknown how broadly they apply to the tasks faced by language learners. Computational models can serve as formal theories of probabilistic learning by instantiating proposals about the learning mechanisms available in early language acquisition.

This chapter surveys computational studies of early language acquisition across the full range of acquisition challenges, from sound-category learning to syntactic rule learning. We describe criteria for evaluating these models on their adequacy as theories of language acquisition. In particular, we propose that models be evaluated on two criteria: their sufficiency—whether they are able to learn aspects of language given appropriate input—and their fidelity—whether they fit the patterns of success and failure shown by human learners. Applying these criteria reveals that there are considerable similarities between many of the most successful models across a range of tasks. Probabilistic learning has often been characterized as implying that learners compute and apply simple descriptive statistics like co-occurrence and conditional probability. We find that simple statistics do not generalize well from task to task. We argue instead that the most successful computational proposals across tasks attempt to infer a parsimonious description of the data in a structured representational vocabulary.

**Chapter 2** argues that social signals of speakers' intentions are at best noisy markers of reference and that in order to take advantage of these signals fully, learners must integrate information across time. This chapter describes a corpus study that takes steps towards quantifying the informativeness of cues that signal speakers' chosen referent, including their eye-gaze, the position of their hands, and the referents of their previous utterances. In particular, we present results based on a hand-annotated corpus of 24 videos of child-caregiver play sessions with children from 6 to 18 months old.

The first parts of this chapter describe our corpus, our coding methods, and the results of our analyses of social cues. Our data suggest that either individually or taken together, social cues are very noisy. Although they carry information about caregivers' referents, even a learner who knew which social cues to attend to would still only be able to guess the speakers' referent correctly a fraction of the time. This result strongly suggests that learners aggregate information about reference across time. In the second part of the chapter, we explore one way that learners might compensate for the noisiness of referential cues: by assuming that reference is continuous and that what was being talked about is quite likely to be the same as what was being talked about in the previous utterance. We provide some intuitions about this assumption of referential continuity and show analyses that suggest that it is a good assumption to make about child-directed speech. We conclude by using a classification analysis to investigate how much information about speakers' referential intentions can jointly be extracted by combining information from social cues with an assumption of discourse continuity. Taken together, these results provide support for a view of early word learning as communicative or intentional inference: the aggregation of information about reference from non-linguistic social cues and from language in service of inferring word meanings.

**Chapter 3** of the thesis describes the kinds of inferences that learners can make by assuming that speakers are informative with respect to their intended meaning, introducing and testing a formalization of how Grice's pragmatic maxims can be used for word learning. This chapter's work begins with the assumption that communicators do not always use language to code facts about the world directly. Instead, speakers and listeners rely on shared assumptions to allow them to communicate more efficiently than if every assumption were made explicit. If word learners take these implicit, shared assumptions into account, they should be able to make

better guesses about what words mean than if they simply assume that language codes true facts about the world. An information-theoretic formulation is introduced to capture the Gricean assumption that speakers choose their words in order to be informative about a target meaning, given some referential context.

Although the basis of our framework is general, making predictions within it requires a model of the space of possible meanings and how they map to natural language expressions. Thus, in order to make a first test of our framework, we study simple games that are similar to the “language games” proposed by Wittgenstein. In the games we study, the shared task of communicators is to identify an object from a set using one or a few words. This very restricted task allows us to define the possible meanings that communicators entertain, in turn allowing us to define an intuitive mapping between words and meanings: that a word stands for the subset of the context it picks out (its extension). Although these simplifications bring our tasks further away from natural language use, they also allow us to derive strong quantitative predictions from our framework. We show how, in these simple language games, our framework derives the *size principle*—a statistical principle found in previous work on word learning and generalization. In Experiment 1, we test whether adults make inferences that are quantitatively congruent with the size principle when they are presented with contextual word-learning inferences like the interior-decorator example above. In Experiment 2, we show that three-year-olds also succeed in making such inferences in a version of the same task. This work takes a first step towards formalizing the pragmatic assumptions necessary for effective communication in under-constrained, real-world situations.

**Chapter 4** presents a fully-realized model of cross-situational, intentional word learning that both learns words and infers speakers' referential intentions from labeled corpus data. This model is based on the idea that cross-situational word learning presents a “chicken-and-egg” inference challenge. If a child could understand speakers' utterances, it would be easy to learn the meanings of individual words; and once a child knows what many words mean, it is easy to infer speakers' intended meanings. To the beginning learner, however, both individual word meanings and speakers' intentions are mysterious. The chapter describes a Bayesian model that solves these two inference problems in parallel, rather than learning exclusively from the inferred meanings of utterances or relying only on cross-situational word-meaning associations. The model infers pairings between words and object concepts from CHILDES data with high precision. The model also explains a variety of behavioral phenomena from the word learning literature, as the result of making probabilistic inferences about speakers' intentions. These phenomena include mutual exclusivity, one-trial learning, cross-situational learning, the role of words in object individuation, and the use of inferred intentions to disambiguate reference.

The thesis proposes a general perspective on the role of statistical inference in language acquisition—that many aspects of language acquisition can be described as statistical inference within a generative model structured around the communicative use of language—and then attempts to formulate and test some aspects of this perspective in the domain of word learning. This work looks forward towards models that describes not only what children *should* learn from the input (sufficiency) but also what they *do* learn (fidelity). A major benefit of the probabilistic framework within which the models in the thesis are formulated is that they can be composed with a wide variety of other models. The broader project of which the thesis forms a part is to formulate models of word learning that go beyond the level of learning names for objects and

address some of the fundamental questions that remain in understanding children's ability to acquire a lexicon.

The contributions of the thesis are fundamentally interdisciplinary, with broad relevance for psychologists, linguists, and computer scientists. First, the thesis reviews the broad literature on language acquisition models from all three fields, focusing on areas of commonality across different models. Second, the thesis is methodologically diverse, including corpus/observational studies, experimental studies with both adults and children, analytic/mathematical work, and statistical modeling. Third, the thesis uses tools from statistics and machine learning to describe phenomena in human cognitive and linguistic development. The word learning model presented in Chapter 4 is applied to a broad range of experimental results from children's word learning, while the framework for informative communication in Chapter 3 is used to make novel quantitative predictions about behavior. Finally, the corpus materials developed in the thesis have already been used in a number of projects in the natural language processing community, pointing to the interdisciplinary interest in word learning.