A much-celebrated aspect of language is the way in which it allows us to express and comprehend an unbounded number of thoughts. This property is made possible because language consists of several combinatorial systems which can be used to creatively build novel words and sentences using large inventory of stored, reusable units.

For any given language, however, there are many more potentially storable units of structure than are actually used in practice --- each giving rise to many ways of forming novel expressions. For example, English contains suffixes which are highly productive and generalizable (e.g., -ness; Lady-Gagaesqueness, pine-scentedness) and suffixes which can only be reused in specific words, and cannot be generalized (e.g., -th; truth, width, warmth). How are such differences in generalizability and reusability represented? What are the basic, stored building blocks at each level of linguistic structure? When is generalization possible when is it not? How can the child acquire these systems of knowledge?

I will discuss how tools from machine learning, artificial intelligence, and computational linguistics can address these problems. The general approach is based on the idea that the problem of computation and storage can be solved by using a probabilistic tradeoff between a pressure to store fewer, more reusable units and a pressure to account for each linguistic expression with as little computation as possible.

This tradeoff is grounded in foundational principles of inductive inference, but has surprisingly far reaching implications across multiple levels of linguistic structure. I will discuss several specific models based on this framework and provide examples of how the approach can help solve long standing empirical puzzles, simplify existing theories, and connect linguistic theories to psychology and computer science.

Bio: Tim is a Research Scientist at MIT and in his research, he develops mathematical models of language generalization, learning, and processing. His research draws on experimental methods from psychology, formal modeling techniques from natural language processing, theoretical tools from linguistics, and problems from all three.

Host: Dr. Steven Piantadosi, spiantadosi@bcs.rochester.edu

Pizza will be provided at 12:15 PM