

Project Summary: “Collaborative Research: An Integrated Theory of Syntactic Acquisition – Realistic input, quantitatively defined target states, and computational models of the learning strategy”

A complete theory of the cognitive architecture that gives rise to the syntactic system of human language will consist (minimally) of three components: (i) a description of the linguistic input available to children (the *input*), (ii) a description of the cognitive mechanisms that underlie adult syntactic behavior (the *target state*), and (iii) a learning mechanism that children use to map between the input and the target state (the *learning strategy*). The field is closer than ever to constructing its first set of large-scale theories of syntactic acquisition because both the computational and experimental methodologies necessary to specify each theoretical component are now available to researchers who are willing to combine their expertise. First, thousands of hours of freely available child-directed speech coming from a variety of sources can be structurally-annotated in order to quantitatively assess children’s input. Second, formal acceptability judgment experiments made available by the field of *experimental syntax* can be used to precisely quantify the gradient target state of learning. Third, computational models of the learning strategy can be designed to reach the target state, given realistic child-directed input. This framework has previously been applied to some constraints on long-distance dependencies (*syntactic islands*); however, extending this methodology to additional syntactic phenomena raises difficult questions about (i) how the learning strategies for each phenomenon can ultimately be integrated into a cohesive theory of syntactic acquisition, and (ii) how the components of those strategies arise in the learner (i.e., are they innately specified, or are they derivable from other components?). We propose to leverage the research framework described above, extending the computational modeling framework to implement learning strategies that can successfully acquire knowledge of several additional phenomena central to syntactic theory from realistic child-directed input. We will focus on the distribution of gaps in multiple gap constructions (parasitic gaps and across-the-board constructions) and the distribution of noun phrases (e.g., Case theory), as these phenomena are likely to highlight the problems inherent in expanding strategies that successfully learn syntactic islands to (i) other gap constructions, and (ii) unrelated syntactic phenomena.

Intellectual Merit: This research project is appealing to both theoretical linguists and the broader psycholinguistic community since it proposes a methodology for creating theories of syntactic acquisition that integrate realistic input data, quantitatively defined gradient target states, and computational models of the learning strategy. Moreover, the project demonstrates how this methodology can be applied to complex linguistic phenomena central to linguistic theory, such as multiple gap long-distance dependencies and the distribution of NPs described by Case theory. From the perspective of language acquisition, this research offers a way to test which learning strategies are capable of solving the induction problems presented by acquisition, and how well those strategies transfer from one linguistic phenomenon to another. From the perspective of linguistic theory, the proposed research examines phenomena that have traditionally been used to motivate Universal Grammar and abstract mechanisms central to modern syntactic theories. This project thus connects theoretical and experimental work in linguistics with computational modeling, yielding results not achievable from each sub-field alone.

Broader Impacts: This project will provide an empirical foundation for studies of syntactic acquisition and the impact of input variability, as well as providing a template for future projects that involve a coordinated synthesis of theoretical syntax, experimental psychology, and computational modeling. The results will be disseminated in multiple ways, including making the annotated corpora and judgment data available to the larger research community and running a workshop designed to stimulate activity across the field in this area. In addition, this project integrates research and education by providing hands-on research experience in theoretical linguistics to multiple undergraduate students and one graduate student each year. It also enhances these students’ discovery and understanding by giving them continued access to professors engaged in highly relevant research that promises to have a significant impact on the field. The collaboration of two PIs from these traditionally disparate sub-fields demonstrates how theoretical linguists and computational modelers can benefit from each other, and thus inspire productive collaboration across the traditional divides of theory, experimentation, and computation.