Productivity and Lexical Specificity in Morphophonological Learning

Gaja Jarosz

Morphophonological learning involves the acquisition of an interactive system of productive generalizations along with morpheme-specific and lexically-specific knowledge. Acquisition of this complex system of knowledge proceeds incrementally on the basis of noisy, incomplete, and ambiguous linguistic input. Due to recent developments in statistical, computational modeling of morphophonological learning, there now exist numerous approaches for learning of various kinds of hidden morphophonological structure from incomplete, unlabeled, and noisy data. These computational models make it possible to connect the full representational richness of phonological theory with noisy, ambiguous corpus data representative of language learners' linguistic experience to make detailed and experimentally testable predictions about language learning and generalization. In this talk, I discuss several ongoing projects that utilize these mutually-informing connections between statistical learning, phonological theory, and experimental data to test hypotheses about the representations that underlie morphophonological and lexical knowledge. These projects seek to bring new sources of evidence to bear on long-standing theoretical debates by differentiating and testing the predictions of distinct theoretical proposals and learning mechanisms. Focusing on several case studies examining the interactions between morphophonological and lexical learning, I demonstrate how statistical learning models can formalize the link between linguistic theory and the distributional properties of the learner’s input. These direct links allow for detailed examination of each theory’s precise predictions for a wide range of empirical phenomena, including the time course of acquisition, generalization to novel items and contexts, as well as online lexical and morphological processing.