Learning Probabilistic Phonology: Integrating Theoretical, Computational, and Developmental Perspectives

Gaja Jarosz

A rapidly growing research area in phonology investigates the grammatical underpinnings of gradient acceptability and the kinds of computational models that can explain human learning of such knowledge from language input. An important and consistent finding of this work is that knowledge of gradient and statistical patterns references the same phonological representations and principles that are relevant in the categorical domain. However, this body of work has largely ignored the question of how knowledge of static restrictions typically assumed to be involved in such tasks relates to the more traditional purview of phonological theory, namely morpho-phonological alternations and processes. Specifically, how does knowledge of gradient phonotactics interact with speakers' phonological grammars and how are these types of knowledge acquired? In this talk I develop an answer to these questions that provides a unified perspective on (gradient) phonotactics and phonology and a formal theory of their acquisition. Building on collective insights from language development, statistical learning theory, and phonological theory, I propose a computational model that casts phonological learning as a statistical inference problem in which learning of stochastic phonotactics serves as a statistical prior. This formalizes the developmental findings indicating that phonotactic learning precedes and informs the learning of phonological alternations. An immediate consequence of this formalization of the relationship between phonotactics and phonology is a general solution to a long-standing problem in language learning known as the subset problem. More generally, this approach makes possible more realistic, more efficient, and more successful learning models that can be used to exploit the full richness of phonological theory in detailed modeling of acquisition. I show that this integrative approach establishes a mutually informing link between linguistic theory and language acquisition. I demonstrate via computational simulation that the proposed interaction between phonotactics and phonology provides a simple explanation for experimental findings on the acquisition of phonological alternations and that it leads to new, testable predictions for language acquisition.