



DISSERTATION DEFENSE



Caleb Belth

Towards an Algorithmic Account of Phonological Rules and Representations

Wednesday, June 21, 2023

10:00am – 12:00pm

3725 Beyster

Hybrid – [Zoom](#)

Passcode: 959010

ABSTRACT: The development of computer science provided a valuable tool for the study of language as a cognitive system, by allowing linguistic theories to be stated in computational terms. These theories have usually placed emphasis on describing the space of possible human languages, and viewed this delineated space as antecedent to a theory of how such a language might be learned from linguistic data. In the domain of phonology—the study of the structure of linguistic sound—this dissertation takes steps approaching the problem from the opposite direction, by framing the problem as that of identifying the learning procedure(s) by which humans construct a language in response to linguistic exposure. The object of study is shifted from the investigation of how a learner will discover a supposed target grammar, to the investigation of the ontogenetic process by which humans develop computational, phonological systems.

The proposed algorithmic approach starts by identifying independently-established psychological mechanisms available to a learner, and then uses these as the components of a hypothesized learning procedure. This dissertation includes a proposed algorithmic account of how abstract representations of words can be constructed, which render long-distance dependencies as local in a graph structure, and allow for effective generalization to unseen words in the face of the sparsity of linguistic input. The dissertation also proposes an algorithmic account of how rules can be constructed to map between these abstract representations and their concrete realizations. Stated in algorithmic terms, the proposed learning system is evaluated on realistic natural language data, and makes precise, testable predictions. The learner constructs accurate linguistic generalizations from training data of no more than a thousand words: across languages evaluated, the learner achieves, on average, 0.96 accuracy on held-out test words, and never lower than 0.92. Moreover, the models' predictions are consistently borne out in developmental predictions and experimental settings, including a novel experiment carried out to directly test this model.

When compared to a prominent alternative learning-based model—neural networks—the proposed model achieves higher accuracy, while producing comparatively interpretable outputs, and—critically—providing an intelligible algorithm, which brings greater understanding to the mechanisms underlying phonological development.

CHAIR: Prof. Danai Koutra and Prof. Andries Coetsee