Bigger data about smaller people: Studying language learning at scale

Michael C. Frank

Abstract: Every typically developing child learns to talk, but children vary tremendously in how and when they do so. What predicts this variability, and what is consistent across children and across learners of different languages? In this talk, I'll describe our efforts to create predictive models of early language learning as a way of formalizing hypotheses in this space. This goal has led us to create open data resources like Wordbank, childes-db, and Peekbank that capture data from tens of thousands of children learning dozens of different languages.