Language as informative imagination

Dr. Gašper Beguš, Department of Linguistics, UC Berkeley

In this talk, I propose a framework for modeling language as informative imagination, offering insights into human and non-human communication as well as into how AI models learn. I introduce ciwaGAN, a model that treats language learning as informative imitation and exhibits several linguistically and neurally motivated properties, including communicative intent, learning from raw unsupervised audio, and embodied representations. I also present an interpretability technique (CDEV), which allows us to perform linguistic and neural simulations on the models. I present a case study in which AI interpretability technique (CDEV) enabled the discovery of analogs to human vowels in whale communication. Understanding how generative models learn--and how they resemble or differ from humans--can not only provide insights into human language and cognition but also facilitate scientific discovery.