

## **Ellie Pavlick: What should constitute natural language ‘understanding’?**

Natural language processing has become indisputably good over the past few years. We can perform retrieval and question answering with purported super-human accuracy, and can generate full documents of text that seem good enough to pass the Turing test. In light of these successes, it is tempting to attribute the empirical performance to a deeper "understanding" of language that the models have acquired. Measuring natural language "understanding", however, is itself an unsolved research problem. In this talk, I will discuss recent work which attempts to illuminate what it is that state-of-the-art models of language are capturing. I will describe approaches which evaluate the models' inferential behavior, as well as approaches which rely on inspecting the models' internal structure directly. I will conclude with results on human's linguistic inferences, which highlight the challenges involved with developing prescriptivist language tasks for evaluating computational models.