

Keynote Speech

Every Time We Hire an LLM, the Reasoning Performance of the Linguists Goes Up

Harish Tayyar Madabushi

University of Bath

Abstract

Pre-Trained Language Models (PLMs), trained on the cloze-like task of masked language modelling, have demonstrated access to a broad range of linguistic information, including both syntax and semantics. Given their access to both syntax and semantics, coupled with their data-driven foundations, which align with usage-based theories, it is valuable and interesting to examine the constructional information they encode. Early work confirmed that these models have access to a substantial amount of constructional information. However, more recent research focusing on the types of constructions PLMs can accurately interpret, and those they find challenging, suggests that an increase in schematicity correlates with a decline in model proficiency. Crucially, schematicity—the extent to which constructional slots are fixed or allow for a range of elements that satisfy a particular semantic role associated with the slot—correlates to the extent of “reasoning” needed to interpret constructions, a task that poses significant challenges for language models. In this talk, I will begin by reviewing the constructional information encoded in both earlier models and more recent large language models. I will explore how these aspects are intertwined with the models’ reasoning abilities and introduce promising new approaches that could integrate theoretical insights from linguistics with practical, data-driven approaches of PLMs.