

Evaluating the Linguistic Competence of Large Language Models: Experimental Evidence from Center-embedding Structures

Mengkai Wang

Beijing Foreign Studies University, Beijing, Beijing, China

Abstract

This study investigates whether large language models (LLMs) possess human-like syntactic competence by examining their handling of English center-embedded structures. Two experiments were conducted: one collected acceptability judgments from native speakers, confirming sensitivity to syntactic constraints; the other measured surprisal values from GPT-2 and Gemma 2 on grammatical versus ungrammatical center-embedded sentences. While both models distinguished between the two types probabilistically, they failed to replicate categorical human judgments. The results suggest that LLMs conflate low-frequency constructions with ungrammaticality, reflecting limitations in hierarchical syntactic understanding. It is argued that genuine linguistic competence requires more than statistical pattern recognition and advocate for integrating formal syntactic theory into model development. This work contributes to the ongoing dialogue between generative linguistics and AI, highlighting key distinctions between human cognition and current LLMs.