

Exploring AI-Generated English Relative Clauses in Comparison to Human Production

Hongoak Yun (Jeju National University)
hongoakyun@jejunu.ac.kr

Human behavioral studies have consistently indicated a preference for subject-extracted relative clauses (SRCs) over object-extracted relative clauses (ORCs) in sentence production and comprehension. Some studies have further shown that this preference can be influenced by the semantic properties of head nouns, particularly animacy. In this study, we use AI language models, specifically GPT-2 and ChatGPT 3.5, to simulate human sentence generation. Our primary goal is to evaluate the extent to which these language models replicate human behaviors in sentence production and identify any divergences. We tasked the models with completing sentence fragments structured as 'the,' followed by a head noun and 'that' (*The reporter that ...*). We varied the semantic property of head nouns such that they are all animate (*the secretary that ...*) in Study 1 and are either animate or inanimate (*the musician/book that ...*) in Study 2. Our findings reveal that in Study 1, both GPT models exhibited a robust SRC bias, replicating human-like behavior in relative clause production. However, in Study 2, we observed divergent behavior between the models when head nouns were inanimate, while consistency was maintained when head nouns were animate. Specifically, ChatGPT 3.5 generated more ORCs than SRCs in the presence of inanimate head nouns. These results, particularly those from ChatGPT 3.5, closely mirror human relative clause production patterns. Our study highlights the potential of language generative language models as efficient and versatile corpus simulators. Furthermore, our findings contribute to the evolving field of AI linguistics, shedding light on the capacity of AI generative systems to emulate human-like linguistic patterns in sentence production.