Pseudo-Psycholinguistic Behavior of Large Language Models

Patrick Juola[1], George Mikros[2], Lise Menn[3], Jean Berko Gleason[4], Nan Bernstein Ratner[5]

[1]Duquesne University – juola@mathcs.duq.edu

[2]Hamad Bin Khalifa University – gmikros@hbku.edu

[3]University of Colorado – lise.menn@colorado.edu

[4]Boston University – gleason@bu.edu

[5]University of Maryland, College Park – nratner@umd.edu

Abstract

Large Language Models (LLMs) have shown human-like linguistic abilities on a variety of tasks. ChatGPT, for example, can answer questions, write essay prompts, edit texts, and even write code. Because of the opacity of the models, though, we do not know how they work, and they are unlikely to use the same processing as humans. In this paper, we use Berko's (1958) Wug Test to study how some existing LLMs respond when required to process English inflectional and derivational morphology. The Wug Test was developed to study whether and how English-acquiring children develop an understanding of the general processes of making new well-formed words from existing words and from word-like novel stimuli ("nonce words"). Children were prompted with spoken sentences containing nonce words (supported by pictures of novel objects and actions) and asked to fill in a pause appropriately; the best known example is "This is a wug. Now there are two of them. There are two ___." We tested ChatGPT-4 on several grammatical patterns from Berko (1958) to see if LLMs can also inflect nonce words (without picture support). Using a sample of roughly 170 nonce stems in noun or verb contexts, we prompted the system to produce the plural, past tense, present progressive, and adjectival forms. Preliminary results show that the system produced nonce plural forms, past tense verbs, possessive markers, and agentive nouns accurately, but could not process the prompt or carry out the task for progressives and adjectives. Some other types of correct forms (compound-like phrases, `fipt house'; past progressive, `was tribbing') were produced sporadically.

Keywords: Large language models, LLMs, Wug Test, psycholinguistics, morphology, nonce words.