Understanding Quantifier Scope with Large Language Models: How Many Children Climbed Trees?

Shaohua Fang
Purdue University, West Lafayette, Indiana, United States

Yue Li
Purdue University, West Lafayette, Indiana, United States

Yan Cong
Purdue University, West Lafayette, Indiana, United States

Abstract

Sentences with multiple quantifiers often lead to interpretive ambiguities, which can vary across languages. This study adopts a cross-linguistic approach to examine how large language models handle quantifier scope interpretation in English and Chinese, using probabilities to assess interpretive likelihood. Humanlikeness scores were used to quantify the extent to which LLMs emulate human performance across language groups. Results reveal that most LLMs prefer SS interpretations, aligning with human tendencies, while only some differentiate between English and Chinese in IS preferences, reflecting human-like patterns. Humanlikeness scores highlight variability in LLMs' approximation of human behavior, but their overall potential is notable. Differences in model architecture, scale, and pretraining data, particularly models' pre-training data language background, significantly influence how closely LLMs approximate human quantifier scope interpretations. Deepseek-R1 was also explored for its potential in handling quantifier scope in English and Chinese.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*