Parallelograms Strike Back: LLMs Generate Better Analogies than People

Qiawen Ella Liu
Princeton University, Princeton, NJ, USA

Raja Marjieh
Princeton University, Princeton, NJ, USA

Jian-Qiao Zhu
Princeton University, Princeton, NJ, USA
The University of Hong Kong, Hong Kong

Adele E. Goldberg
Princeton University, Princeton, NJ, USA

Thomas L. Griffiths
Princeton University, Princeton, NJ, USA

Abstract

Four-term word analogies (A:B::C:D) are classically modeled geometrically as "parallelograms," yet recent work suggests this model poorly captures how humans produce analogies, with simple local-similarity heuristics often providing a better account (Peterson et al., 2020). But does the parallelogram model fail because it is a bad model of analogical relations, or because people are not very good at generating relation-preserving analogies? We compared human and large language model (LLM) analogy completions on the same set of analogy problems from Peterson et al. (2020). We find that LLM-generated analogies are reliably judged as better than human-generated ones, and are also more closely aligned with the parallelogram structure in a distributional embedding space (GloVe). Crucially, we show that the improvement over human analogies was driven by greater parallelogram alignment and reduced reliance on accessible words rather than enhanced sensitivity to local similarity. Moreover, the LLM advantage is driven not by uniformly superior responses by LLMs, but by humans producing a long tail of weak completions: when only modal (most frequent) responses by both systems are compared, the LLM advantage disappears. However, greater parallelogram alignment and lower word frequency continue to predict which LLM completions are rated higher than those of humans. Overall, these results suggest that the parallelogram model is not a poor account of word analogy. Rather, humans may often fail to *produce* completions that satisfy this relational constraint, whereas LLMs do so more consistently.

Keywords: Analogy; LLM; Parallelogram model

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*