Human and LLM processing of scope ambiguities in English ditransitives

Existing experimental research has shown that both humans and large language models (LLMs) are able to derive linear and non-linear scope in simple SVO scopally ambiguous sentences. Despite this, little experimental work on the processing of scope ambiguities has focused on structures in which particular scopal readings are impeded by syntactic structure, i.e., in island constructions, or those in which scope is frozen, i.e., the English double object construction (DOC). Judgments of such unavailable readings are robust amongst linguists and trained speakers cross-linguistically, but experimental evidence from naive speaker judgments are often inconsistent with theoretical proposals. Further, to my knowledge, these structures have remained unexplored in the linguistic ability of LLMs. Ditransitives are of particular interest as it has been argued that prepositional dative constructions (PDCs) are rightward ascending, while DOCs are rightward descending, meaning surface scope is non-linear in PDCs but linear in DOCs:

Figure 1: PDC versus DOC structure

(a) Rightward ascending (PDC) **(b)** Rightward descending (DOC)

Not only do we predict a strong preference for linear scope compared to non-linear scope under this theoretical framework due to the frozen scope effect, but also a preference for non-linear scope over linear scope in PDCs.

To assess the theoretical predictions outlined above, I present a 2x2x2x2 acceptability judgment study in which participants rate continuations to scope ambiguous sentences compatible with either linear or non-linear scope. The conditions manipulated are island presence (present vs. not present), ditransitive structure (PDC vs. DOC), indirect object quantifier type ($\forall$ vs. $\exists$), and continuation type (singular vs. plural NP). The experimental design thus results in 16 unique sentence-pair conditions in which three conditions (two islands and one frozen scope DOC), according to the theory discussed, are not predicted to be available. 320 native speakers of British English rated 20 experimental, and 40 filler, sentence pairs on a 7-point Likert scale. This paradigm is duplicated with minor adjustments to test GPT-2 (all sizes—small, medium, large, and xl) and

Llama 3.1 (@8B parameters). I extract log probabilities values for continuation NPs. To render human and LLM results comparable, Likert scale ratings and log probabilities are transformed into surprisal.

The results confirm previous findings that both humans and LLMs can derive linear and non-linear readings in scope ambiguous sentences in which both readings are predicted to be available. However, it also appears that neither naive speakers nor LLMs are sensitive to the frozen scope effect. While not unexpected for LLMs, this finding provides further support which raises doubts of the reality of the frozen scope among untrained speakers, as no study to my knowledge has succeeded in detecting a lack of non-linear scope in English DOCs.

Additionally, the results provide evidence that English PDCs are rightward ascending as naive speakers more readily derive non-linear scope readings of PDCs than linear readings. This effect is not replicated in LLMs, suggesting that LLMs solely take linear order into account in deriving scope preferences, while humans are sensitive to hierarchical structure when resolving scope ambiguous data.