Understanding perspectival language is important for applications like dialogue systems and human-robot interaction. We present a dataset for evaluating perspective inference in English, ProSPer, and use it to explore how humans and Transformer-based language models infer perspective.

**Key contributions:**
- ProSPer: a novel dataset for probing understanding of spatial perspectival language.
- Novel human behavioral data showing that humans achieve around 77-88% accuracy.
- Comparison of neural language models, showing that RoBERTa’s accuracy is human-like.
- Fine-grained error analysis guided by previous psycholinguistic work, revealing a genre frequency bias for humans and RoBERTa.

**Predicting Spatial Perspective Requires:**
- Determining who is important enough to be a perspective-holder (Grosz et al. 1995)
- Gathering and evaluating contextual evidence
- Resolving ambiguity
- Inferring spatial relations

**ProSPer: Probing Spatial Perspective**

**Task:** given a passage with an omitted verb, decide if the missing word is come or go.

**Example:** Rick changed the subject. “I heard that you were having some furniture delivered this afternoon,” he said to Aunt Emily. “I thought I’d ___ by and see if you needed any help.”

(1) go  (2) come

**Automatically selected subset:** 47385 examples of come, go, walk, drive, and arrive from the OANC

**Annotated subset:** 600 examples from Davies, 2008, 2016, 2011 annotated for perspective-holder, destination, syntactic environment, and tense.

**Human Performance**
- Human judgments collected on 3 ProSPer subsets:
  - Random: 600 items randomly sampled from the Automatic subset
  - NN Confounding: the 300 Automatic items most challenging for NN models.
  - Annotated: the entire Annotated subset
- 300 participants recruited through Prolific
- Target verb presented with its semantic competitor
- Bidirectional context provided

**Neural Network Performance**

Evidence tentatively supports a genre frequency bias: RoBERTa is best at predicting perspective with syntactic environments and perspective-holders common in text; humans do better in conversational contexts.

**Summary**

**Exploring Perspectival Biases**

**Strong Egocentricity Hypothesis**
- Low accuracy for come relative to other verbs (Epley et al., 2004; Lin et al., 2010).

**Weak Egocentricity Hypothesis**
- High accuracy with speaker perspectives (Harris, 2012; Anderson, 2020).

**Genre Frequency Bias Hypothesis**
- Human accuracy improved by conversation-like contexts and speaker or listener perspectives.
- RoBERTa accuracy improved by text-like contexts and perspectives common in text.

**Citations**