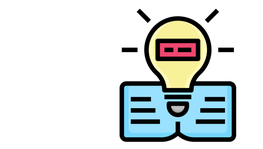
Meaning Beyond Truth Conditions
Evaluating Discourve Level Undestanding via Anaphora Accessibilities

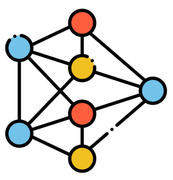
Xiaomeng Zhu* Zhenghao Zhou* Simon Charlow Robert Frank
Department of Linguistics, Yale University

TL;DR

- We present a hierarchy of natural language understanding abilities and argue for the importance of the **discourse level.**
- We propose the task of **anaphora accessibility** and develop an evaluation dataset inspired by theoretical research in dynamic semantics.
- We find that LLMs and humans align on some tasks and diverge on others, and such divergence can be explained by LLMs' reliance on specific lexical items as opposed to abstract structure as indicators of discourse structure.



Dynamic Semantics Theories



Large Language Models

Acceptability Judgments

Existential vs. Universal

Explicit / Lexical ∀

A farmer worked in his field. He dreamed of the harvest.

Every farmer worked in his field. # He dreamed of the harvest.

Infelicitous because outside of scope

■ Discourse entity ■ **Quantifier** scope ■ Anaphora

$\exists x(\text{Farmer}(x) \wedge \text{WorkInField}(x)) \wedge \text{Dream}(x)$

Implicit / Embedded ∀

The farmer owns a donkey, and he beats it. It is a big one.

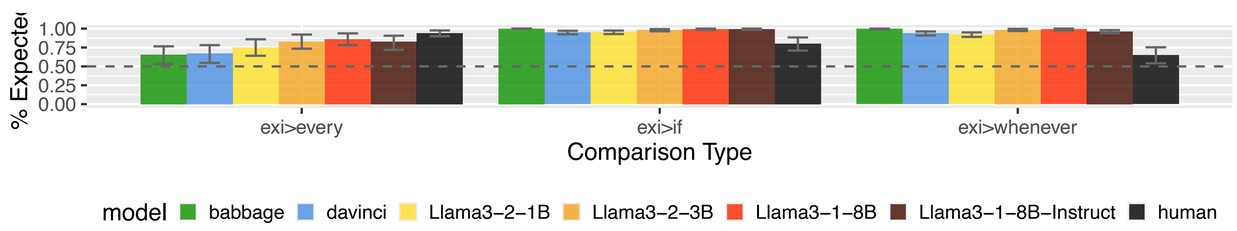
If the farmer owns a donkey, he beats it. # It is a big one.

Infelicitous because outside of scope

■ Discourse entity ■ **Quantifier** scope ■ Anaphora

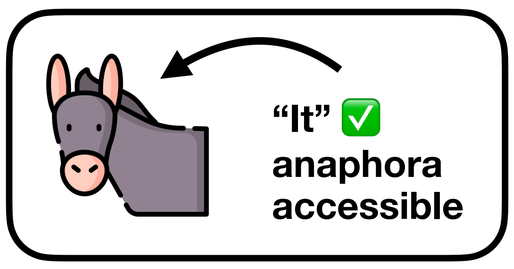
$\forall x[\text{Farmer}(x) \rightarrow \text{WorkInField}(x)] \wedge \text{Dream}(x)$

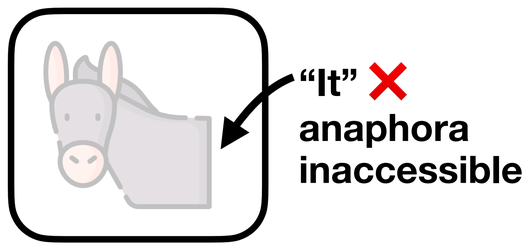
Prediction: $P(cont \mid \exists) > P(cont \mid \forall)$



Conclusion: LLMs are sensitive to the difference in anaphora accessibility between universal and existential quantifiers.

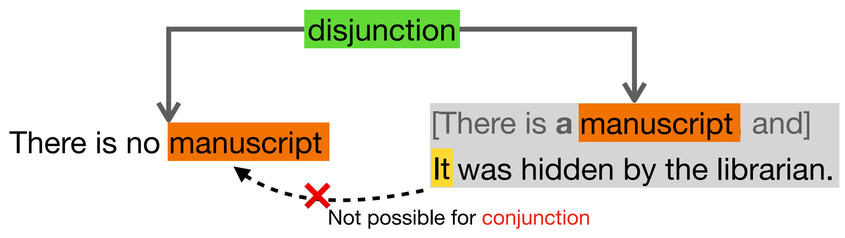
Sentence boundaries:
∀, ∃, ¬, ∨, ∧





Anaphora Accessibility

Disjunction vs. Conjunction

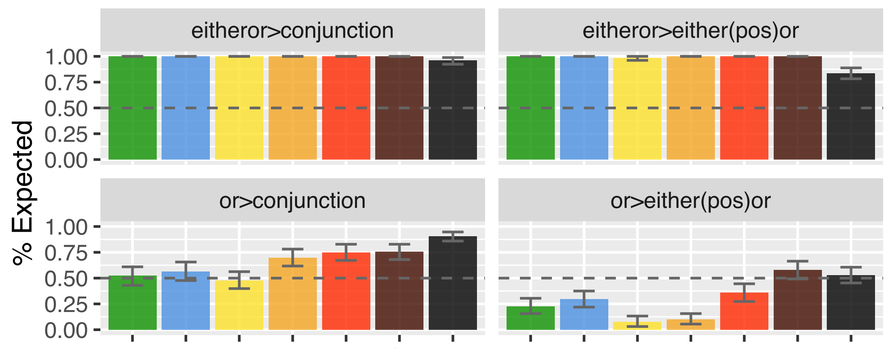


■ Discourse entity ■ Scope ■ Anaphora

$[\neg \exists x(\text{Manuscript}(x))] \vee [\exists x(\text{Manuscript}(x) \wedge \text{Hidden}(x))]$

$\neg \exists x(\text{Manuscript}(x)) \wedge \text{Hidden}(x)$

Prediction: $SLOR(disjunction) > SLOR(conjunction)$



Conclusion: LLMs are sensitive to the difference between disjunction and conjunction, but are also subject to lexical factors.

Negation

The farmer owned a cow. It was away on the meadow.

It was not the case that the farmer didn't own a cow.

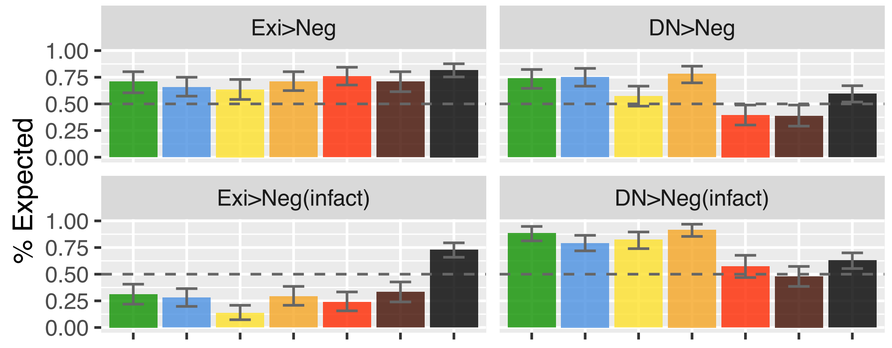
The farmer didn't own a cow. # It was away on the meadow.

■ Discourse entity ■ Scope ■ Anaphora

$\exists x(\text{Cow}(x) \wedge \text{Own}(\text{farmer}, x)) \wedge \text{Away}(x)$

$\neg[\exists x(\text{Cow}(x) \wedge \text{Own}(\text{farmer}, x))] \wedge \text{Away}(x)$

Prediction: $P(cont \mid \exists, \neg\neg) > P(cont \mid \neg)$

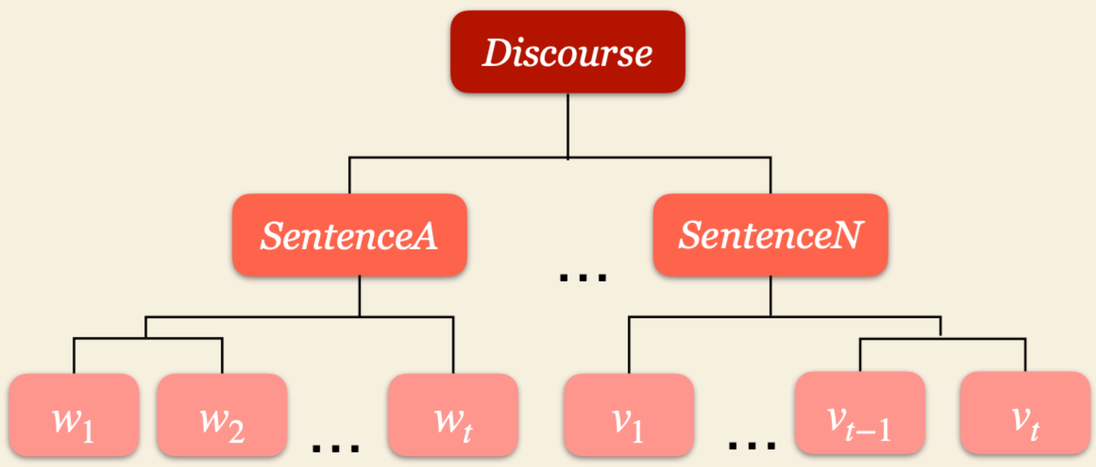


Conclusion: LLMs recognize that both Exi and DN are better than Neg, but are influenced by lexical factors.

Hierarchy of NLU

- Level 3: Discourse Level Understanding: *Integration of sentence meaning to coherent discourse meaning*
- Level 2: Sentence Level Understanding: *Integration of lexical meaning to coherent sentential meaning*
- Level 1: Lexical Level Understanding: *a word's extension and intension*

Natural Language Structure



Example Tasks

- Long-Document Understanding; Discourse Entity Recognition & Tracking; Discourse Parsing…
- Natural Language Inference; Pragmatic Inference; Scope Ambiguity…
- Word Sense Disambiguation; Taxonomy Construction; Lexical Entailment; Analogical Relation…

ACL 2025 Main Conference @ Vienna, Austria

{miranda.zhu, herbert.zhou, simon.charlow, robert.frank}@yale.edu