Ambiguity

Austin Hester
CS 5300 – NLP Presentation
April 4, 2024

What is Ambiguity?

- An **ambiguity** – a failure of the speaker to clarify their intended meaning from alternative interpretations.
  - The speaker's intended meaning of an utterance is unclear to the listener who is consciously aware of an ambiguity.
  - Almost every sentence is ambiguous in some way.

- **Lexical** ambiguity –a word has multiple definitions.

- **Syntactic** ambiguity – a phrase has multiple grammatical parses.

- **Semantic** ambiguity – a phrase has multiple meanings.
  - Naturally follows lexical and syntactic ambiguities.

BERT Shows Garden Path Effects

Tovah Irwin, Kyra Wilson, Alec Marantz.

Problem

Garden path sentence – a sentence which misleads with ambiguity at first until a syntactic inconsistency forces the reader to reconsider their interpretation. *Temporal ambiguity.*

The horse raced past the barn fell.

While I was surfing the internet went down.

The boat floated down the river sank.

Motivation

- Ambiguity is unavoidable in natural language.

- Garden path sentences provide insight into the processing of ambiguous structures.
  - Their self-correcting structures provide lessons on sentence parsing.

- Humans have difficulty parsing these sentences, so the researchers wanted to know if transformer models, such as BERT, also face the same challenge.

Q&A Implementation

- Comparison of role assignments between humans and four BERT-style models in a question answering task.
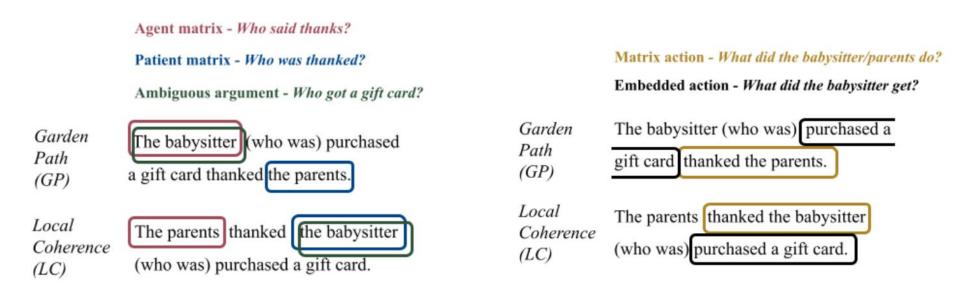 - Standard and garden path sentences included.

Agent matrix - *Who said thanks?*

Patient matrix - *Who was thanked?*

Ambiguous argument - *Who got a gift card?*

Garden Path (GP)
The babysitter (who was) purchased a gift card thanked the parents.

Local Coherence (LC)
The parents thanked the babysitter (who was) purchased a gift card.

Figure 1: Example questions and answers targeting correct entity identification.

Matrix action - *What did the babysitter/parents do?*

Embedded action - *What did the babysitter get?*

Garden Path (GP)
The babysitter (who was) purchased a gift card thanked the parents.

Local Coherence (LC)
The parents thanked the babysitter (who was) purchased a gift card.

Figure 2: Example questions and answers targeting correct action identification.

Q&A Results

- Humans and BERT models both show a decline in performance when faced with a question containing a "garden path".

- Humans perform better at semantically connecting words further away from each other.

- Humans often guess when they don't know; models restate the question.

Probe Implementation

- An internal probe uncovers the underlying semantic representations assigned by BERT for each question.

- This probe gives the researchers insight into the representations held at each hidden layer of the models' neural networks.

Probe Results

- The probe confirms that garden path sentences cause semantic confusion in BERT-style models.

- The different models assigned different representations for the same sentences, despite being fine-tuned with the same training data.
  - This shows that garden path sentences are inherently hard to parse regardless of language model.

Solution Quality

- **Sample Size**
  - Fair at 4 models, <100 humans, and 100s of questions.

- **Study Design**
  - Compares Q&A performance between humans and BERT-style models.
  - Analyzes the semantic representations of the Qs as deciphered by the mode

- **Effect Size**
  - Humans 14.9% better at locally coherent sentences.
  - Models between 8.8% to 20.5% (average 15.5%) better at locally coherent sentences.
  - No effect emerges.

Solution Quality

- **Data Analysis Methods**
  - Only the highest-performing layer from each model used for analysis in the probe task, which adds some bias.

- **Consistency of Results**
  - Consistent:
    - Demonstrating difficulty in parsing garden path sentences.
  - Inconsistent:
    - Internal semantic representations between BERT-style models.
    - Performance on certain types of questions between humans and models.
    - The layers at which each model performed their best.

- **Researchers' Interpretation**
  - No significant statistical results found.
  - Similarity in performance between all subjects surprising.

Thank you!

Austin Hester