What I Did for 224U Final Project

Systematicity in GPT-3's Interpretation of Novel English Noun Compounds

Siyan Li, Riley Carlson, Christopher Potts

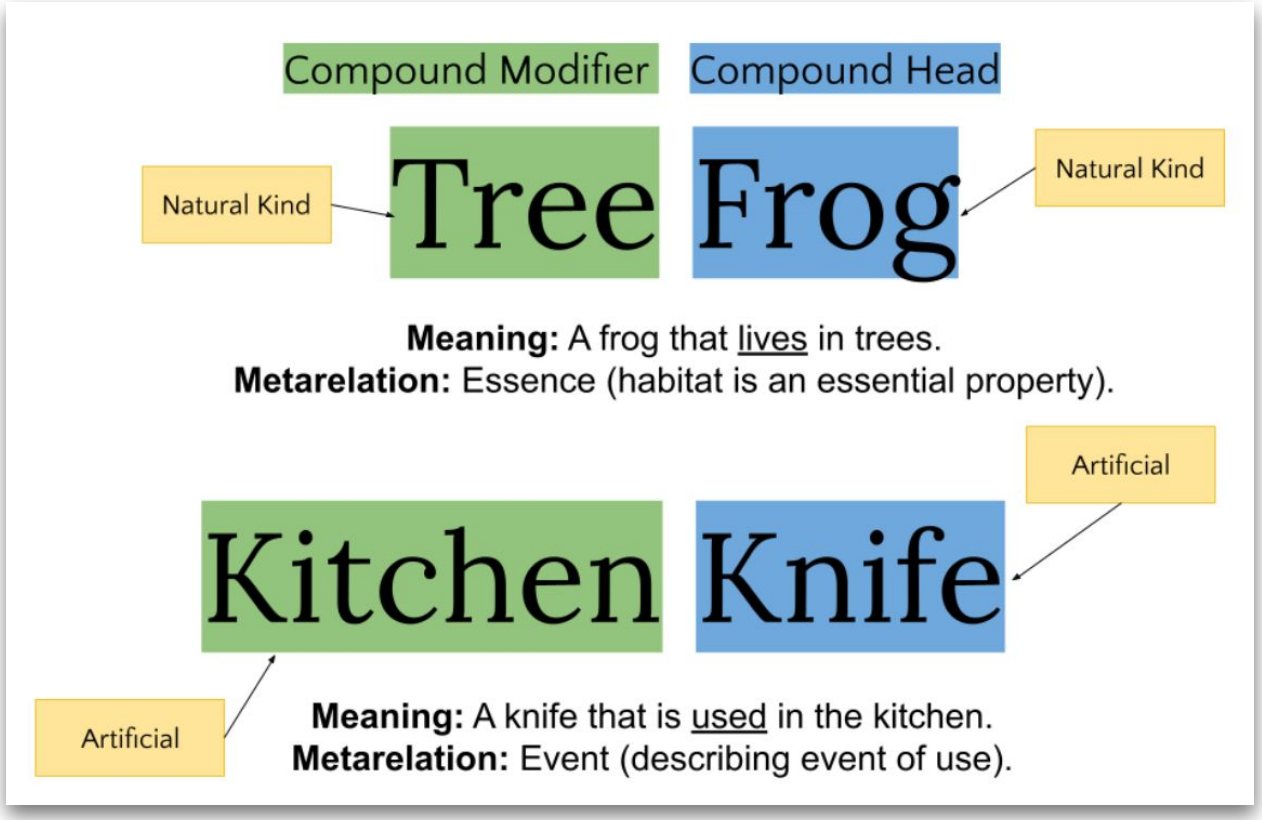
1

Motivation for Project

Why is this project important?

Inspirational Paper

- Levin et al. *Systematicity in the semantics of noun compounds: The role of artifacts vs. natural kinds.*
- Asked human participants to provide explanations of novel noun compounds (e.g. stew skillet, duck screen)
- **Events vs. Essences Hypothesis:**
  - The modifier in an **artifact-headed compound** typically refers to an *event of use or creation* associated with that artifact;
  - The modifier in a **natural kind-headed compound** typically makes reference to *inherent properties reflective of an abstract essence* associated with the kind, such as its perceptual properties or native habitat.
- I know this looks overwhelming but I will explain in the next slide


Compound Modifier
Compound Head
Natural Kind
Tree
Frog
Natural Kind
Meaning: A frog that lives in trees.
Metarelation: Essence (habitat is an essential property).
Artificial
Kitchen
Knife
Artificial
Meaning: A knife that is used in the kitchen.
Metarelation: Event (describing event of use).

Despite the compounds being novel, the human intuition for their meanings remains systematic

Another Motivation

All these language models are great, but we have no idea what is going on. Which begs the question:

Does GPT-3 (at the time of this paper) have systematicity in its interpretation of noun compounds?

How I did the Lit Review

- A literature review looks different based on what inspired your work
  - Is it an extension of an existing paper for a different model or approach? Explain that existing paper in more detail than the rest
  - Your literature review should give enough information for **anyone who is in the field of NLP / your target audience** to keep up with your work

How I did the Lit Review

- Usually you have to decompose your research question, and research each facet and write sections accordingly
  - But **no need to go too low-level**! You don't get into the nitty-gritty until the methods / experiments
  - For this work, we study compounds via using GPT-3 to perform language generation

2 Background

2.1 English Noun Compounds

2.2 LLMs and Linguistic Creativity

2

Experimental Protocol

How do we find out?

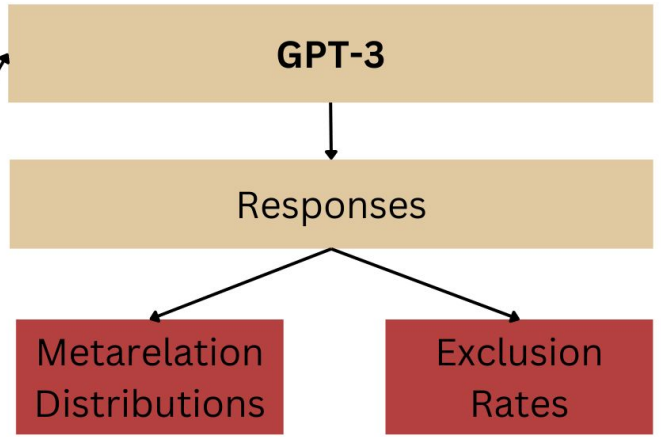
Natural: Imagine that you encounter the compound X. What would you think this refers to?

Structured: Compound: X.\n\nExplanation:

Few-shot

Compound: X_1.\n\nExplanation: E_1\n\n
Compound: X_2.\n\nExplanation: E_2\n\n
Compound: X_3.\n\nExplanation: E_3\n\n
Compound: X.\n\nExplanation:

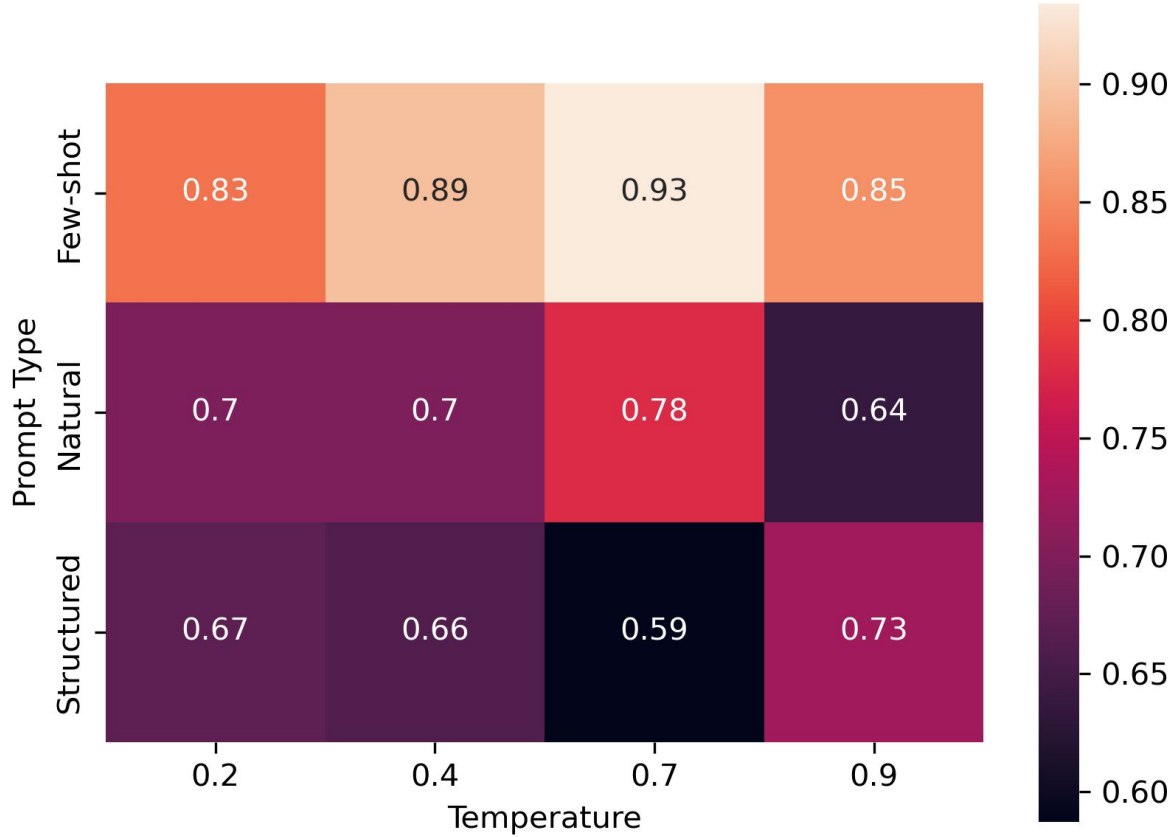
Prompts



Pipeline for all experiments

Experiment 1: Levin et al. Novel Compounds

- We ***directly use the 38 novel compounds*** from Levin et al., and prompt GPT-3 (text-davinci-002)
- Three annotators, with > 0.7 Fleiss' Kappa
  - ***Side note:*** *If you have human annotators, high inter-rater agreement is important!*
- Compute:
  - Correlation between GPT-3 generation labels & original Levin et al. labels
  - Exclusion rates for each condition



Correlations between GPT-3 & Human labels from Levin et al.

	Temperature			
	0.2	**0.4**	**0.7**	**0.9**
Natural	13.15	10.52	21.05	28.94
Structured	28.94	36.84	28.94	34.21
Few-shot	0.0	0.0	5.26	2.63
Human		17.67		

Exclusion rates of GPT-3 generations in Exp 1

GPT-3 Displays good systematicity so far!

… But does this extend to more novel compounds?

Experiment 2: Even more Novel Compounds

- **Additional novel compounds** with no lexical overlaps!
  - With more varied metarelation subtypes as well
- Compute exclusion rates only for this experiment, since we don't have any "human gold labels" available

	Temperature			
	0.2	**0.4**	**0.7**	**0.9**
Natural	5.56	11.43	5.56	5.88
Structured	14.29	17.14	19.44	44.12
Few-shot	5.71	5.71	11.43	8.57

Okay, fine, so far it looks like GPT-3's behavior is governed by the Events vs. Essences hypothesis...

But we can't rule out that it is just reasoning about lexical items!

Experiment 3: In-context Conceptual Reasoning (The Good Stuff)

- We use complete nonsense strings to eliminate lexical clues
  - gmtomflxri and putrlv
- We hypothesize that:
  - GPT-3 **will flop** this test
  - The **type of compounds** in the prompt would **influence** the **performance**
- We pre-select a set of noun compounds, then construct compounds with different semantic distances & different head and modifier types based on the selection

A modifier is definition$_1$.
A head is definition$_2$.
A compound is definition$_3$.\n\n
A gmtomflxri is definition$_4$.
A putrlv is definition$_5$.
A gmtomflxri putrlv is

Noun	Strawberry
Match-close (Mc)	Banana
Match-far (Mf)	Bamboo
Different-close (Dc)	Shortcake
Different-far (Df)	Overcoat

Assuming that **Strawberry Cookie** is the basis compound, and that we select **Banana Table** to be the new compound

A strawberry is sweet fleshy red fruit.
A cookie is any of various small flat sweet cakes.
A strawberry cookie is a cookie made with strawberries.\n\n
A gmtomflxri is a tropical yellow fruit.
A putrlv is a piece of furniture.
A gmtomflxri putrlv is

Prompt Structure and Example

Head	Mc	Mf	Dc	Df	Avg.
Modifier					
Mc	25.0	50.0	0.0	75.0	37.5
Mf	0.0	0.0	0.0	0.0	0.0
Dc	50.0	25.0	75.0	33.3	45.8
Df	75.0	25.0	50.0	50.0	50.0
Avg.	37.5	25	31.25	39.5	

Exclusion Rates from Exp 3

It is likely that GPT-3 is reasoning about lexical items, since the performance here is much worse!

Example: “Biden Burrito”

The Biden burrito would most likely refer to a burrito that is made in the style of Vice President Joe Biden. This could mean that the burrito is filled with Biden’s favorite foods, or that it is made to look like Biden himself.

The most likely interpretation of “Biden burrito” would be a burrito named after United States Vice President Joe Biden.

Burrito that is eaten by Biden.

The Biden burrito is a political term used to describe a 2020 United States presidential election campaign event in which then-Presidential candidate Joe Biden ate a burrito on a live stream.

Thanks!

Any questions ?