

What Mechanisms Underlie Linguistic Generalizations In Large Language Models?

Investigate How LLMs Process Generics and Noun-noun Compounds

Giulia Rambelli, Ph.D.

May 2, 2025 - GroNLP reading group



ABSTRACTION Research Group

Adele Loia Marianna M. Bolognesi Pre-doc P Linguistics, Clinical Linguistics and related Linguistic, Semantics disciplines **ABSTRACTION** Giulia Rambelli Tommaso Lamarra Ph.D Post-doc Computational Linguistics, Linguistics, Cognitive Psychology, Cognitive Modeling, Analogy Categorization, Iconicity Caterina Villani Andrea Amelio Ravelli Post-doc Post-doc Computational Linguistics, Cognitive Psychology, Abstract Computer Science, Machine concepts, Categorization, Social interaction, Pragmatics Learning, Multimodality

Generalization is crucial for tasks like understanding complex language structures, solving new problems, and making inferences in real-world situations.[1]



[1] Hammer, B., Ilievski, F., Saralajew, S., & van Harmelen, F. (2024). Generalization by People and Machines

Generalizations in NLP

• **Out-of-distribution generalization**: the ability of a model to accurately classify text instances from distributions different from those of the training data_[2].

- OOD generalization includes:
 - Ability to solve examples from unseen domains (relying on prior knowledge)
 - Robustness to prediction when introducing superficial changes to the input
 - Updating the prediction when introducing semantic changes to the input

[2] Yang, J. et al. (2023). Out-of-distribution generalization in natural language processing: Past, present, and future. In EMNLP 2023(pp. 4533-4559).

Linguistic Generalizations

- Linguistic generalization refers to the ability to **apply known language rules or structures to novel examples**.
 - Example: If you know how to form plural nouns (e.g., "cat" → "cats"), you can generalize this rule to unfamiliar words like "dog" → "dogs".

• LLMs have demonstrated the ability to generalize in many tasks (e.g., translation, grammar prediction), but challenges remain in abstract reasoning.

Today's Talk

• Study 1 – Generic Sentences Interpretation

- Provides new insights into how sensitive LLMs are to generic statements.
- Study 2 Noun-noun Compound Interpretation
 - Explores the ability of LLMs to interpret both lexicalized and novel compounds.

Generics (or generic generalizations)

Dogs are mammals

Birds fly

Ducks lay eggs

Mosquitoes carry malaria

7

- Statements that express generalizations about a category.
 - The truth of a generic is not related to prevalence of the property
 - Generic expressions tolerate exceptions, and in various degrees
 - → Generics ≠ Universally quantified generalizations

GENERIC EXPRESSION	QUANTIFICATION VALUE
Dogs are mammals	All dogs are mammals
Birds fly	<i>Most</i> birds fly (but not all!)
Ducks lay eggs	Less than half of ducks lay eggs (only adult females)
Mosquitoes carry malaria	Some mosquitoes carry malaria (less than 1%)

Generics (or generic generalizations)

- The interpretation of generic expressions involves pragmatic reasoning and world knowledge_[3]
- Psycholinguistic experiments on the human interpretation of generics
 - \rightarrow mainly based on contrasting generics with overtly quantified sentences.



Research Questions

Implicit Quantificational Value of Generics
 RQ1: Can LLMs interpret generics based on semantic content?



2. Default Interpretations of Generic Form **RQ2:** Do LLMs apply a default interpretation to generics?

ACL 2024 Collacciani, C., Rambelli, G., & Bolognesi, M. (2024). Quantifying generalizations: Barghak Thailard Exploring the divide between human and llms' sensitivity to quantification. ACL 2024

RQ1: Implicit Quantificational Value of Generics

• Task 1: Surprisal Distributions

• Probabilities of quantifiers are the same independently of the sentence meaning

Task 2: Prompting

- Preference for *some* when given only quantified sentences.
- Preference for the generic sentence (inherently vague) when included.

"Most lions have manes." *"Few* lions have manes."

Tell me which of the following is the most truthful sentence: Birds fly.

No birds fly. Few birds fly. Some birds fly. Most birds fly. All birds fly.

 Overall: LLMs are not particularly sensitive to the meaning of quantifiers and are not able to discriminate between the different sentences actual quantificational meanings as humans do.

Models: Llama2-7b and Mistral-7b (base+instruct)

RQ2: Default Interpretation of Generic Form

• When informed through a generic about novel categories, humans tend to estimate percentages on average very close to 100 percent (= *all*).

• Prompt the models to see if they replicate this **default interpretation of generics**.

Task from Cimpian et al. (2010)

Information:

Morseths have silver fur.

<u>Question</u>:

What percentage of morseths do you think have silver fur?

Our prompting task

```
What is the correct completion? Birds
fly, therefore...
no birds fly.
few birds fly.
some birds fly.
most birds fly.
all birds fly.
```

Tell me which of the following is the most truthful sentence:

No birds fly. Few birds fly. Some birds fly. Most birds fly. All birds fly.





What is the correct completion? Birds fly, therefore... no birds fly. few birds fly. some birds fly. most birds fly. all birds fly.

100 -

80

60

40

20



Findings

- LLMs do not show a strong sensitivity to quantifiers.
 - LLMs do not have the commonsense knowledge required to interpret generic and quantified sentences with respect to their semantic content.
- LLMs seem to encode a meaning associated with the generic form.
 - LLMs skew their preferences toward «strong» positive quantifiers (*most* and *all*), if the generic is provided as a premise in an entailment condition.
 - This default interpretation of generics mirrors that of humans.

Implications

• Generics are a default mechanism for generalization

- Children acquire generics earlier than explicitly quantified statements_[4]
- Generics allow people to reason with incomplete information and draw inferences in novel situations [5]
- LLMs tend to overextend generic statements to most/all members of realworld categories.
 - Such behavior may contribute to stereotype encoding—a potentially harmful bias.
 - Further research is needed to understand how LLMs interpret generics and how this affects their overall behavior when looking at social categories.

[4] Leslie, S.J., & Gelman, S.A. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. Cognitive Psychology, 64(3):186–214
 [5] Asher, N. & Morreau, M. (1995). What some generic sentences mean. *The Generic Book*, 300–338.

Today's Talk

• Study 1 – Generics Interpretation

• Provides new insights into how sensitive LLMs are to generic statements.

Study 2 – Noun-noun Compound Interpretation

• Explores the ability of LLMs to interpret both lexicalized and novel compounds.

Conceptual Combination or Compounding

Two lexical concepts are often used together as phrases to represent a combined concept of greater specificity_[6].



Combining words is a hallmark of language generativity, or **productivity**.

[6] Dunbar & Myers (1988). Concept combination and the characterization of lexical concepts. In Hüllen, Werner; Schulze, Rainer (eds.). Understanding the lexicon: meaning, sense and world knowledge in lexical semantics, 292–302.

16

AVOCADO CHAIR



Productivity through Analogy

- Interpreting a novel compound involves:
 - accessing the **concepts** denoted by the words
 - selecting a **relation** to form a unified conceptual representation
- Hypothesis from Gagné and colleagues_[7,8]
 - The on-line interpretative processing of novel nominal compounds is affected by **analogous lexicalized compounds**

mud man

milk man 'A man who delivers mud' *garbage man* 'A man who collects mud'

[7] Gagné & Shoben (2002). Priming Relations in Ambiguous Noun-noun Combinations. *Memory & Cognition*.
 [8] Gagné & Spalding (2006). Conceptual Combination: Implications for the Mental Lexicon. *The Representation and Processing of Compound Words*.

Conceptual Combinations in LLMs

 LLMs are somewhat able to generalize to new examples that are conceptually similar to training examples. [9]

ChatGPT 3.5 ~

You

Provide a possible definition of "emergency butter".

ChatGPT

"Emergency butter" refers to a reserve supply of butter kept on hand for unexpected or urgent situations, such as running out of butter during meal preparation or when unexpected guests arrive. This backup butter ensures that cooking or baking can continue without interruption and helps avoid any last-minute trips to the store.

4) 1 5 7

• ..But they *still* over-rely on similar training examples, and are brittle when these examples are manipulated. [10]

[9] Coil, J. & Shwartz, V. (2023). From chocolate bunny to chocolate crocodile: Do Language Models Understand Noun Compounds? Findings of ACL 2023.

[10] Shapira, N. et al. (2023). Clever hans or neural theory of mind? stress testing social reasoning in large language models. *arXiv preprint arXiv:2305.14763*.

£

Research Questions

- Interpretating Lexicalized Noun-Noun Compounds (NCs)
 RQ1: Do LLMs Grasp Semantic Relations in Lexicalized Noun Compounds?
- 2. Interpretating Novel Noun-Noun Compounds (NCs)

RQ2: Are LLMs able to generalize semantic relations over novel compounds?



Rambelli, G., Chersoni, E., Collacciani, C., & Bolognesi, M. (2024). Can Large Language Models Interpret Noun-Noun Compounds? A Linguistically-Motivated Study on Lexicalized and Novel Compounds. ACL 2024)

RQ1: Interpreting Lexicalized NC

Task LLMs choose the correct interpretation among 9 paraphrases (to avoid "parroting").

• Use Surprisal and metalinguistic prompting ("which is the most likely description of..?")

• Surprisal Results

- COMP(OSITION)-R and PRODUCTION are almost perfect
- PURPOSE, PROD-R, and TOPIC-R are mostly mistaken
- Compounds characterized by higher **concreteness** are interpreted more accurately (concreteness effect)

compound	Hatcher-Bourque (Pepper, 2022)	paraphrase (Pepper, 2021)
plastic bag	COMP(OSITION)-R	a bag that is composed of plastic
trash bag supermarket shelf	CONT(AINMENT)-R LOCATION	a bag that contains trash a shelf that is located in a supermarket
car door	PARTONOMY	a door that is part of a car
food company	PRODUCTION	a company that produces food
bank loan	PROD(UCTION)-R	a loan that a bank produces
research group	PURPOSE	a group intended for research
art class wind turbine	TOPIC-R US(A)G(E)-R	a class that is about art a turbine that uses wind

668 compositional noun-noun compounds (NC)

RQ2: Interpreting Novel NC

- People generalize the implicit semantic relation of a conventional compound and transfer it to a semantically similar but novel compound.
- We substitute the head/modifier with a hypernym from **WordNet** (64 compounds).

EQUIPMENT BOX

GLOVE BOX

GLOVE COMPARTMENT

- Overall:
 - Changing the modifier is less problematic than changing the head
 - Suboptimal solution: choose PURPOSE relation

- "a container intended for gloves"
- × "a container *that contains* gloves"

RQ2: Interpreting Novel NC

- Interestingly, the same pattern is observed between L1 and L2 English speakers (Italian high school students)
 - Preliminary study: 126 compounds (63 lexicalized + 63 novel), balanced across 7 semantic relations
- Finding: L1 speakers prefer *intended to* similar to LLMs for novel compounds
 - *boat trip* \rightarrow a trip <u>that uses</u> a boat BUT
 - *conveyance trip* \rightarrow a trip <u>intended for</u> conveyance

Results collected for a Master thesis by Marta Mulazzani

"Interpretazione dei composti nominali in inglese: confronto tra apprendenti italiani L2 e nativi inglesi"

Findings

- Performance on Lexicalized NCs:
 - LLMs perform better on compounds with **concrete** concepts.
 - Struggle with abstract or specific relations.
- Generalization Limitations:
 - LLMs struggle with novel compounds, especially when the **head is changed**
 - Difficulty in applying learned relations to novel compounds.
- We demonstrated LLMs' limitations in abstraction and linguistic analogization.

Implications

- Models that grasp abstract concepts and linguistic subtleties can engage in more meaningful conversations with humans.
- The extent to which LMs produce and interpret combinations of constructions offer a deeper understanding of the skills required for natural language.
- Generalizing language structures leads to more accurate answers and predictions in tasks like translation, summarization, and dialogue generation.

Conclusions

- Generalization is a key component of human language and cognition.
- Understanding the underlying mechanisms of linguistic generalization can inform the design of future AI systems, pushing them closer to human-level comprehension and reasoning.







Unlocking meaning from experience through language

ERC-2021-STG-101039777 JUNE 2022- MAY 2027







@abstractionerc.bsky.social

Funded by the European Union (GRANT AGREEMENT: ERC-2021-STG-101039777). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.