

Compositionality and Generalization

Introduction

Shane Steinert-Threlkeld

Mar 25, 2024

Today's Plan

- Introductions
- Motivation / background
 - Compositionality: statement
 - Arguments
 - The homomorphic conception
 - Generalization
 - Guiding themes and questions
- 15 minute break
- Course information / logistics

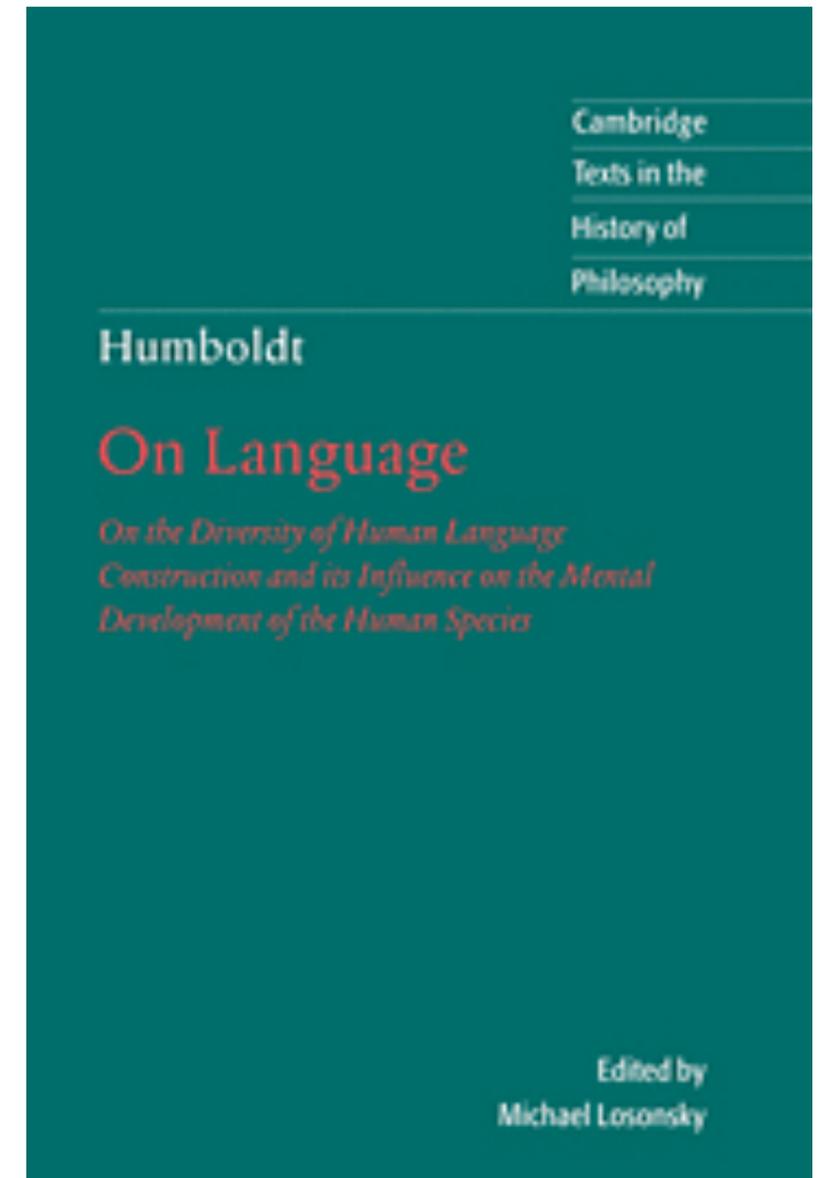
Introductions

- Name
- Program at UW
- What most intrigues you about the topic of compositionality?
- What's one fun thing you did during this past week of spring break?

Motivation: what is compositionality?

Compositionality: towards a statement

- von Humboldt, 1836:
 - “a singing creature... only associating thoughts with tones”
 - “making infinite use of finite means”



Compositionality: towards a statement

VOL. LXXII. No. 285]

[January, 1963

M I N D

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPOUND THOUGHTS

BY GOTTLOB FREGE

[*Translator's Note* :

This article, entitled 'Logische Untersuchungen. Dritter Teil : Gedankengefüge', was published in the *Beiträge zur Philosophie des deutschen Idealismus*, III (1923), 36-51. The first two parts of these 'logical investigations' were 'Der Gedanke' (*Beiträge I* (1918), trans. 'The Thought' in *MIND*, lxv (1956)) and 'Die Verneinung' (*Beiträge I* (1919), trans. 'Negation' in Geach and Black (ed.) *Translations from the Philosophical Writings of Gottlob Frege*). The translation is by Mr. R. H. Stoothoff (St. Antony's College, Oxford), who wishes to thank Mr. P. L. Heath for checking the translation.]

Compositionality: towards a statement

- Frege (1923/1963): “It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a human being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new.”

VOL. LXXII. No. 285]

[January, 1963

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPOUND THOUGHTS

BY GOTTLÖB FREGE

[*Translator's Note* :

This article, entitled ‘Logische Untersuchungen. Dritter Teil : Gedankengefüge’, was published in the *Beiträge zur Philosophie des deutschen Idealismus*, III (1923), 36-51. The first two parts of these ‘logical investigations’ were ‘Der Gedanke’ (*Beiträge I* (1918), trans. ‘The Thought’ in *MIND*, lxxv (1956)) and ‘Die Verneinung’ (*Beiträge I* (1919), trans. ‘Negation’ in Geach and Black (ed.) *Translations from the Philosophical Writings of Gottlob Frege*). The translation is by Mr. R. H. Stoothoff (St. Antony’s College, Oxford), who wishes to thank Mr. P. L. Heath for checking the translation.]

Compositionality: towards a statement

- Frege (1923/1963): “It is astonishing what language can do. With a few syllables it can express an incalculable number of thoughts, so that even a thought grasped by a human being for the very first time can be put into a form of words which will be understood by someone to whom the thought is entirely new.”
- “This would be impossible, were we not able to distinguish parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence serves as an image of the structure of the thought.”

VOL. LXXII. No. 285]

[January, 1963

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPOUND THOUGHTS

BY GOTTLÖB FREGE

[*Translator's Note* :

This article, entitled ‘Logische Untersuchungen. Dritter Teil : Gedankengefüge’, was published in the *Beiträge zur Philosophie des deutschen Idealismus*, III (1923), 36-51. The first two parts of these ‘logical investigations’ were ‘Der Gedanke’ (*Beiträge I* (1918), trans. ‘The Thought’ in *MIND*, lxxv (1956)) and ‘Die Verneinung’ (*Beiträge I* (1919), trans. ‘Negation’ in Geach and Black (ed.) *Translations from the Philosophical Writings of Gottlob Frege*). The translation is by Mr. R. H. Stoothoff (St. Antony’s College, Oxford), who wishes to thank Mr. P. L. Heath for checking the translation.]

Initial Statement

- Rough / general statement (e.g. [Szabó 2022](#)):
- The meaning of a complex expression is determined by its structure and the meanings of its constituents.

The General Picture

- How do we know what “Shane organized a new seminar this quarter.” means?

Arguments for Compositionality

Learnability and Productivity

- Competent speakers of a language can understand—that is, recover the meaning of—*new/novel complex expressions* (unboundedly many).
- For example: the sentence above, let's call it “(S)”.
- Since you have not been exposed to that complex expression ((S)) before, you must be determining its meaning from something you do already know.
- The structure of ((S)) and the meanings of its parts would suffice to determine its meaning.
- So, its meaning is determined by the parts and how they are put together.

Systematicity

- If you know the meanings of certain complex expressions, you automatically know the meanings of certain others.
 - From: “UW is in Seattle”, and “The cat is in the living room.”
 - To: “The cat is in Seattle”, “UW is in the living room”, ...
- The only way that this could be true is if the meanings of the complexes are determined by the parts and how they are put together.

Arguments Against

Arguments Against

- Idioms: “kick the bucket”, “spill the beans”, ...
 - Rejoinders: these are isolated exceptions, can be seen as lexical items

Arguments Against

- Idioms: “kick the bucket”, “spill the beans”, ...
 - Rejoinders: these are isolated exceptions, can be seen as lexical items
- Anaphora (due to Barbara Partee):
 - I dropped ten marbles and found all but one of them. It is probably under the sofa.
 - I dropped ten marbles and found nine of them. # It is probably under the sofa.
 - Rejoinders: (i) change the conception of meaning (e.g. dynamic semantics), (ii) The anaphora really is possible in the second case.

Formal Statement

The Homomorphism Formulation

- Syntax forms an *algebra* $\langle E, \{F_i\} \rangle$
 - Expressions
 - Syntactic formation operations (i.e. “tree-builders”)
- $\mu : E \rightarrow M$ is a meaning-assignment function
- **Compositionality:** for every k -place syntactic operation F_i , there is a k -place semantic operation G_i , such that

$$\mu \left(F_i (e_1, \dots, e_k) \right) = G_i \left(\mu (e_1), \dots, \mu (e_k) \right)$$

(I.e.: μ is a homomorphism from the syntactic to the semantic algebra.)

The Homomorphism Formulation

- Example: “a new seminar”
- NP(Det(a), AdjP(Adj(new), N(seminar)))

Generalization

Generalization

- Productivity and systematicity arguments can be seen through the lens of *generalization*
- A learner is said to generalize from items of type A to items of type B if, when trained on A, it can successfully perform on B.
- (These terms are all intentionally vague.)
- “Compositional generalization”: extending a form-meaning mapping from “simpler” cases to “more complex” cases

A Standard Measure of Generalization

- Train set: examples to train a learner
- Test set: examples to evaluate a learner
 - (In practice, usually also have a validation set, i.e. a three-way split, but not relevant for present purposes.)
- “I.I.D. Generalization”: the training and test set are two *disjoint samples* that come from the *same underlying distribution*
 - (Independent, identically distributed)
 - E.g.: gather a bunch of data, randomly split samples into two sets

Generalization \neq Solving a Task

- One lesson we are learning from recent years with NLP models:
 - Good I.I.D. generalization \neq “solving the task”
 - When the task involves language understanding, we often think that requires compositional generalization.

Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference

R. Thomas McCoy,¹ Ellie Pavlick,² & Tal Linzen¹

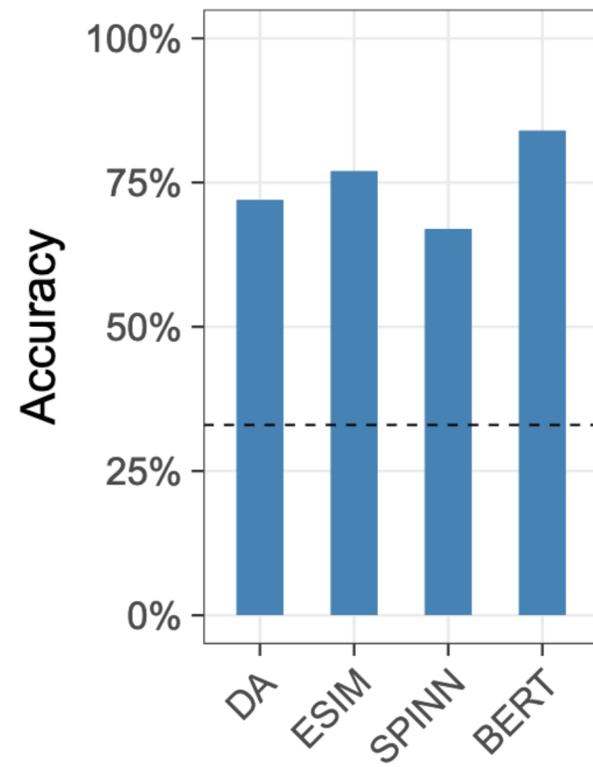
¹Department of Cognitive Science, Johns Hopkins University

²Department of Computer Science, Brown University

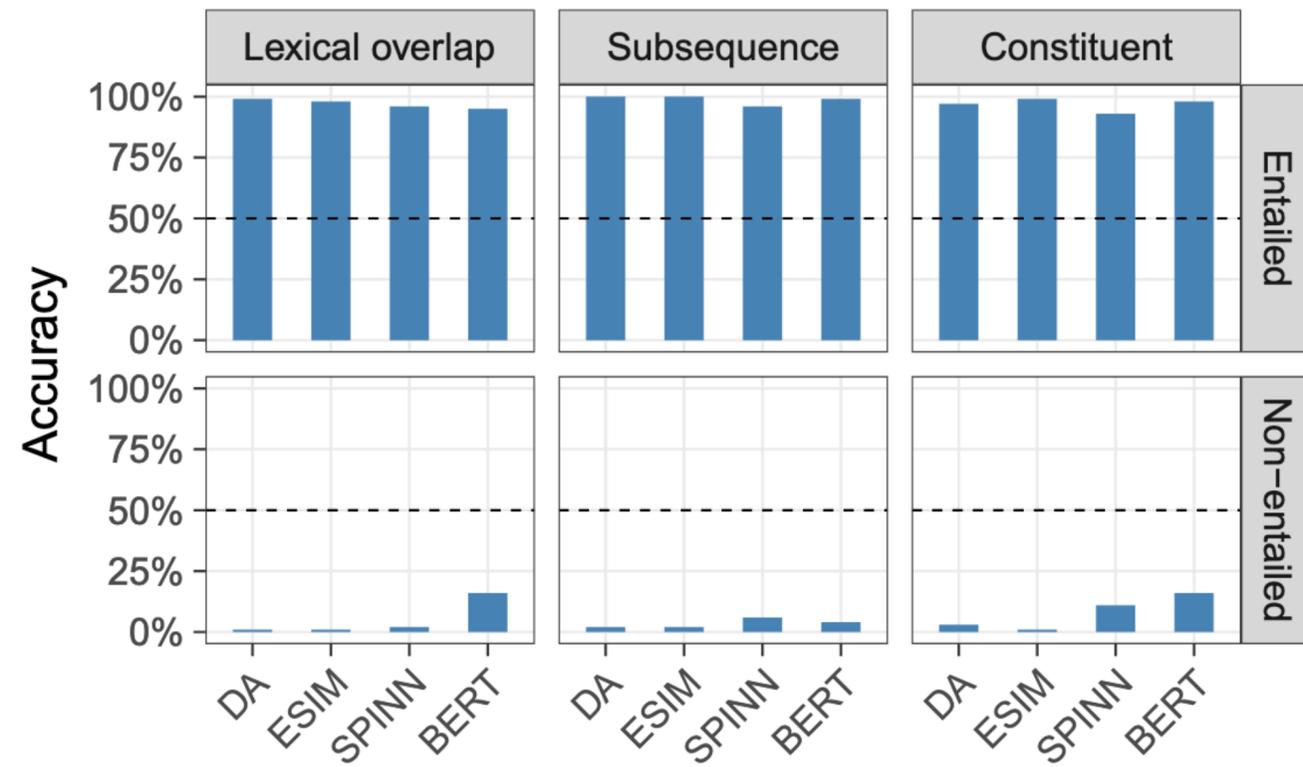
tom.mccoy@jhu.edu, ellie_pavlick@brown.edu, tal.linzen@jhu.edu

Heuristic	Premise	Hypothesis	Label
Lexical overlap heuristic	The banker near the judge saw the actor.	The banker saw the actor.	E
	The lawyer was advised by the actor.	The actor advised the lawyer.	E
	The doctors visited the lawyer.	The lawyer visited the doctors.	N
	The judge by the actor stopped the banker.	The banker stopped the actor.	N
Subsequence heuristic	The artist and the student called the judge.	The student called the judge.	E
	Angry tourists helped the lawyer.	Tourists helped the lawyer.	E
	The judges heard the actors resigned.	The judges heard the actors.	N
	The senator near the lawyer danced.	The lawyer danced.	N
Constituent heuristic	Before the actor slept, the senator ran.	The actor slept.	E
	The lawyer knew that the judges shouted.	The judges shouted.	E
	If the actor slept, the judge saw the artist.	The actor slept.	N
	The lawyers resigned, or the artist slept.	The artist slept.	N

Results

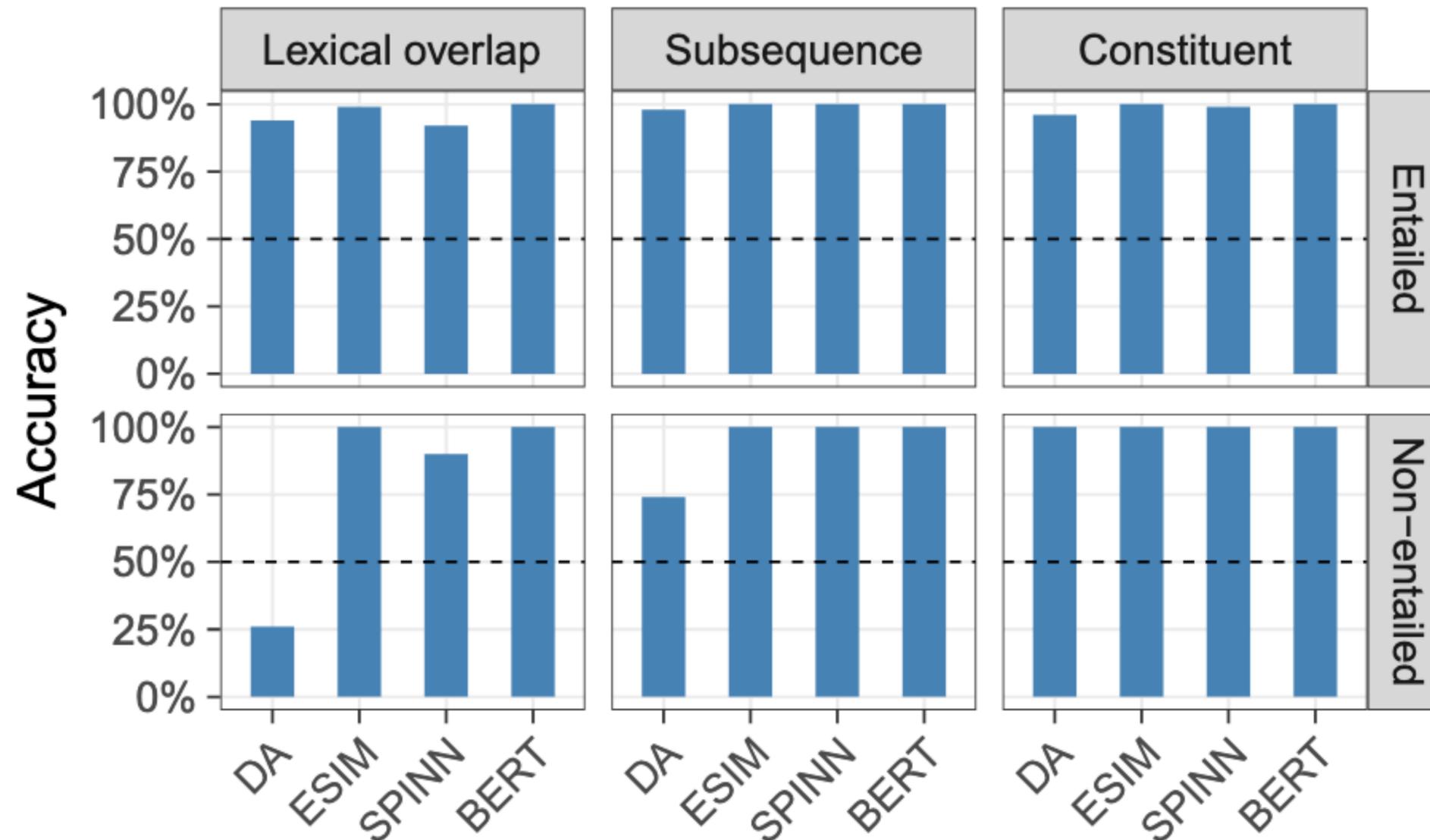


(a)



(b)

Fine-tuning augmented with examples



Measuring and Narrowing the Compositionality Gap in Language Models

Ofir Press^{1,2} Muru Zhang¹ Sewon Min^{1,3} Ludwig Schmidt^{1,4} Noah A. Smith^{1,4} Mike Lewis³

¹Paul G. Allen School of Computer Science & Engineering, University of Washington

²MosaicML ³Meta AI Research ⁴Allen Institute for AI

ofirp@cs.washington.edu

Abstract

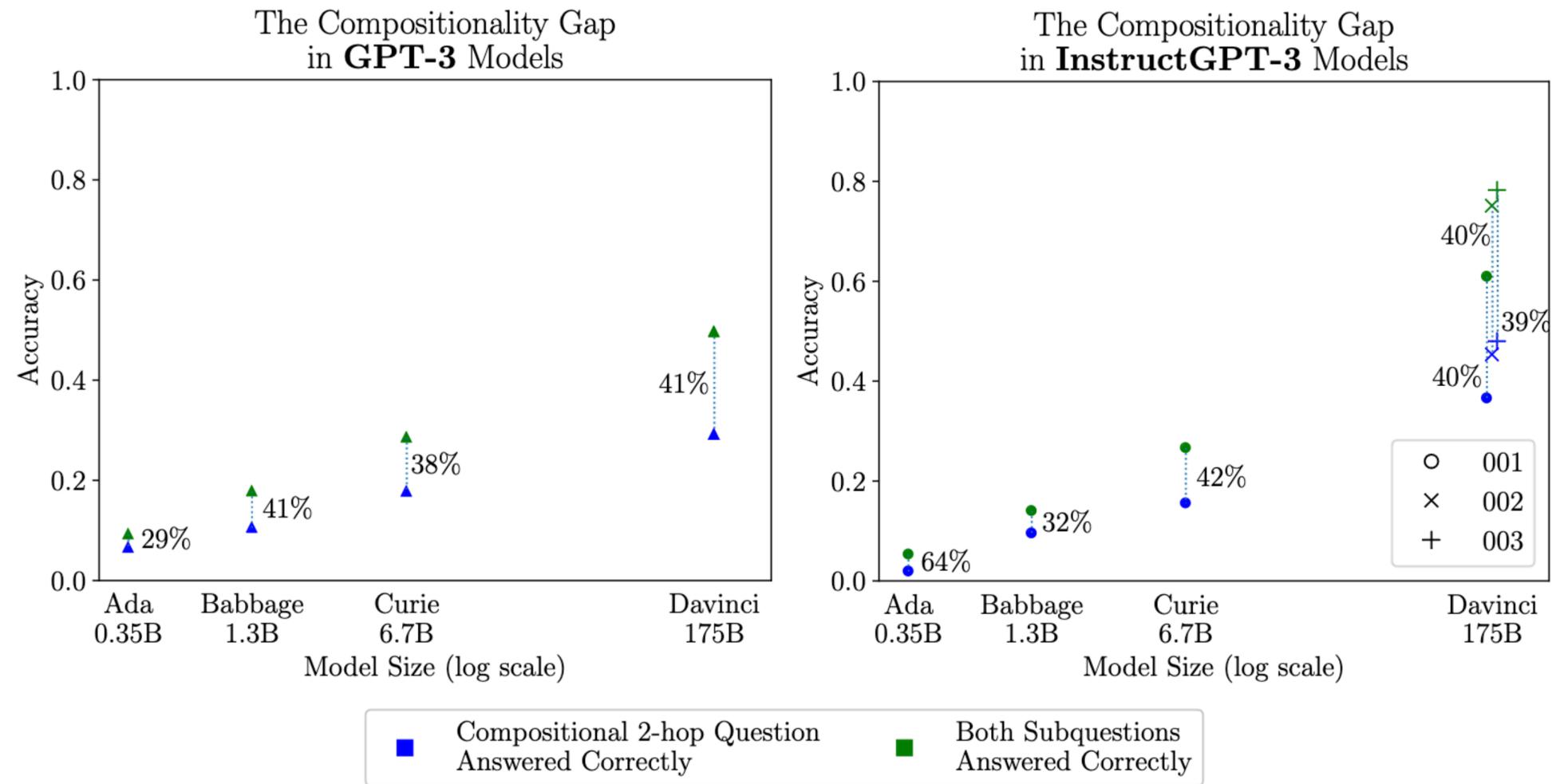
We investigate the ability of language models to perform compositional reasoning tasks where the overall solution depends on correctly composing the answers to sub-problems. We measure how often models can correctly answer all sub-problems but not generate the overall solution, a ratio we call the *compositionality gap*.

soning over these facts. While language models (LMs) have shown strong question answering performance, it remains unclear how much is due to memorization of huge corpora vs. how much is due to reasoning.

First, we quantify the reasoning abilities of LMs using multi-hop question answering. We present a new, automatically generated dataset, Composi-

A Minimal Test Setup

- What city is the University of Washington located in?
- Who is the mayor of Seattle?
- Who is the mayor of the city where the University of Washington is located?



A Taxonomy of Generalization

nature machine intelligence



Analysis

<https://doi.org/10.1038/s42256-023-00729-y>

A taxonomy and review of generalization research in NLP

Received: 22 December 2022

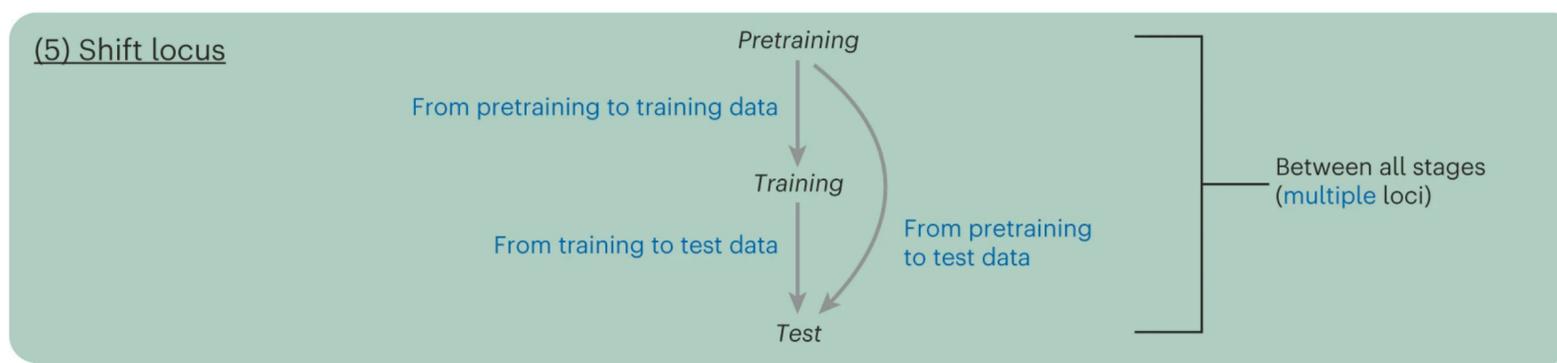
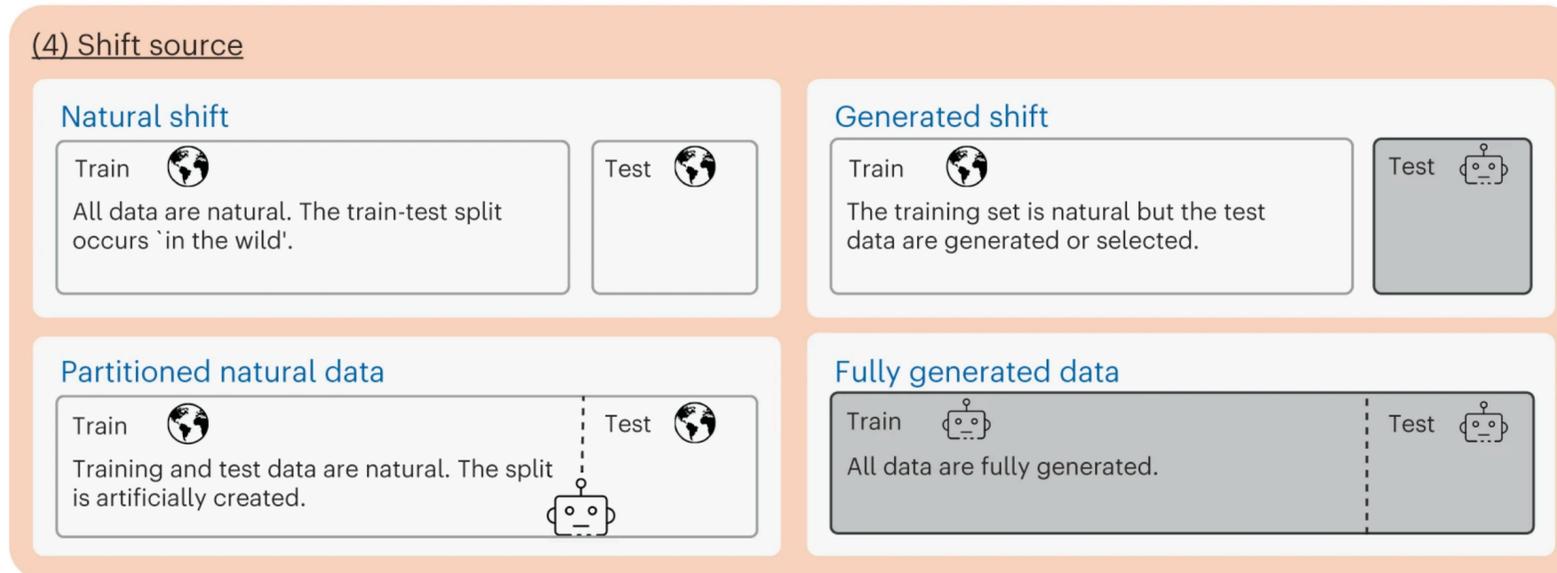
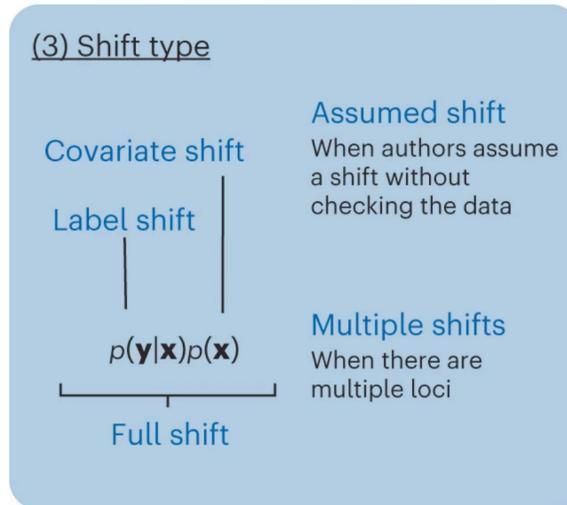
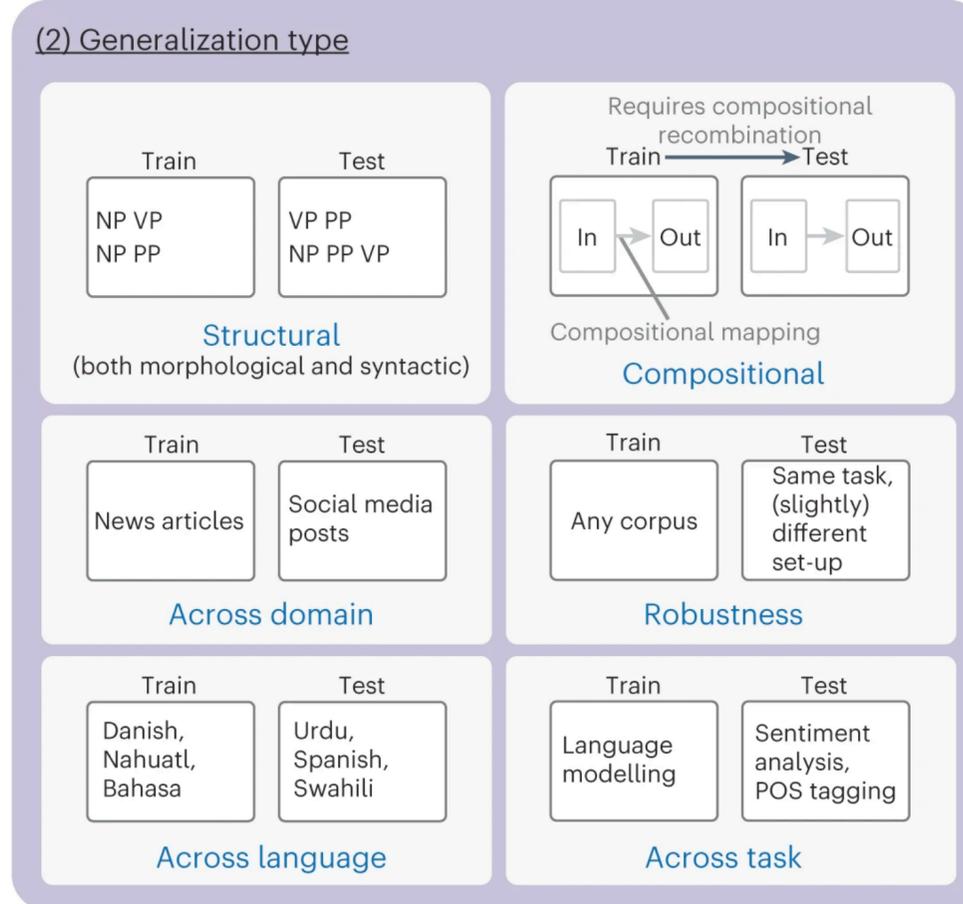
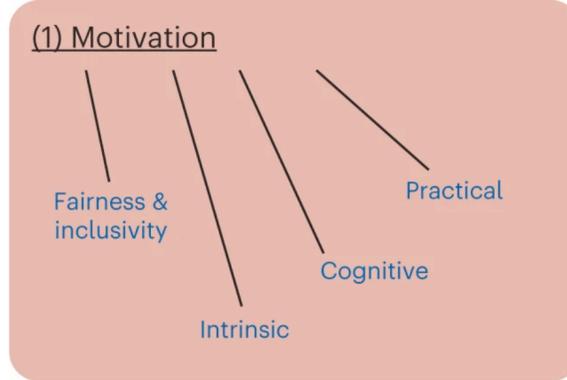
Accepted: 5 September 2023

Published online: 19 October 2023

Check for updates

Dieuwke Hupkes¹✉, Mario Giulianelli²✉, Verna Dankers³✉, Mikel Artetxe⁴, Yanai Elazar^{5,6}, Tiago Pimentel⁷✉, Christos Christodoulopoulos⁸✉, Karim Lasri⁹, Naomi Saphra¹⁰, Arabella Sinclair¹¹, Dennis Ulmer^{12,13}, Florian Schottmann^{14,15}, Khuyagbaatar Batsuren¹⁶✉, Kaiser Sun¹⁷, Koustuv Sinha¹⁷, Leila Khalatbari¹⁸, Maria Ryskina¹⁹✉, Rita Frieske¹⁸✉, Ryan Cotterell¹⁴ & Zhijing Jin^{14,20}✉

The ability to generalize well is one of the primary desiderata for models of natural language processing (NLP), but what ‘good generalization’ entails and how it should be evaluated is not well understood. In this Analysis we



Guiding Themes and Questions

Some questions to keep in mind

- How do we measure the compositionality of a given system?
 - What if the representations are quite unlike what we are used to in formal syntax and logical semantics?
 - Examples: artificial (neural) agents communicating, animal communication systems, ...
- Do we need strictly compositional generalization (or will other forms work) to overcome the I.I.D. problem?
- What might cause failures of compositional generalization?
- How may we build systems that exhibit more human-like generalization?

Stretch Break!

Course Overview / Logistics

Large Scale

- Motivating questions:
 - How can we detect whether a given system does or does not exhibit compositional behavior?
 - How can we build artificial systems that generalize in human-like ways?
- *A research seminar*: in groups, you will carry out and execute a novel analysis project.
 - Think of it as a proto-conference-paper, or the seed of a conference paper.
- *A role-playing reading group seminar*.

Course structure

- Each week, we will read and discuss two papers as a group.
- Role-playing seminar (new format!):
 - Each student will sign up for a role for **one** of the papers, and will be expected to contribute to the discussion from that perspective.
 - Every role must have at least one student, but some roles can have more than one.
 - (For the other paper, you will have to submit a possible alternative title and one discussion question.)
- Sign-up sheet will be shared on Canvas.

The Four Roles

-  **Archaeologist:** Determine where this paper sits in the wider landscape of previous and subsequent work. More concretely: find and report on one prior paper that substantially influenced the current paper and one newer paper that cites this current paper.
-  **Reproducer:** You are attempting to reproduce the main results of the paper, either to confirm its conclusions, or to carry out similar experiments. If you are unable to reproduce the results, try to understand why. If you are able to, explain what helped the most.
-  **Researcher:** Based on the current paper, propose an experiment or analysis that would extend the work in a new direction. This could be a new dataset, a new model, a new evaluation metric, etc. In particular, try to think of something that is only possible thanks to the current paper, not just a simple extension.
-  **Reviewer:** Write a referee report for the current paper, following the form for ACL Rolling Review (up through the "Overall Assessment"). Aim to approach papers charitably, but critically, thinking of how the paper can be made into the best version of itself. Their reviewer tutorial may also be helpful.

Evaluation

- Weekly reading assignments: 60%
 - (Asynchronous option TBA very soon)
- Final project: 40%

Final Project

- More guidelines, and intermediate milestones, to follow.
- Two standard / default options:
 - Reproduce an experiment from a paper we discuss in the seminar.
 - Develop a new compositionality/generalization benchmark, in the format of a submission to the [GenBench Collaborative Benchmarking Task](#)
- Otherwise: open-ended research project focused on the topics of this course. There will be plenty of time to discuss ideas with me as well, including a proposal process in the middle of the quarter.

Reading List

- Relatively broad list of papers on website
 - NB: curated, not exhaustive; impossible to keep up with an entire literature
- Browse, get ideas/inspiration
- Deep dive on a few later
- We can also modify the weekly readings later in the term based on student desire / interest

Current Planned topics

- Measuring compositionality
- Benchmarks I, Benchmarks II
- Architectural approaches
- Other approaches (data augmentation, meta-learning)
- Prompting approaches
- Emergent communication
- Animal communication

Group Formation (HW1)

Three Tasks

- Form groups (more next)
- Set up repository
 - GitHub, GitLab, patas Git server ...
 - Make it private for now!
 - Don't put private or sensitive data in the repo! (incl LDC corpora)
- Add ACL paper template to repository
 - <https://github.com/acl-org/acl-style-files>
 - Format for final paper

Groups

- Groups sized 2-4 people
- Unified grade
- Group decides how to divide work, but reports who did what at the end.
- Aim to diversify talents / interests in the group.
 - Experimental design
 - Data work
 - Implementation
 - Experiment running / analyzing
 - Writing
 - Speaking (presentations)

Communication

- CLMS Student Slack/Discord
 - Useful, since a majority of students in this seminar are on it already
 - Self-organize (575 channel?), based on interests, background competences, etc
 - For students not on it yet:
 - Canvas thread for requesting access
 - CLMS students: please add ASAP
- For general / non-group discussions, still use Canvas discussions.
- NB: I am not on that Slack (nor are other faculty)

Registering Groups

- On Canvas, upload “readme.pdf” with:
 - Group member names, screenshot of repository

Next Week

Measuring Compositionality

- Two main papers:
 - [Measuring Compositionality in Representation Learning](#)
 - [Measuring Compositional Generalization: A Comprehensive Method on Realistic Data](#)
- Sign up for roles here: <https://docs.google.com/spreadsheets/d/1Ine7AR0H-PgsOIQIHZ8YuUmPgZBCuIKJKSGyfkq3PY/edit#gid=0>
- Course website: <https://www.shane.st/teaching/575/spr24/>

Thanks! Looking forward to a great quarter!