Dissociating Formal and Functional Linguistic Competence in Large Language Models

Anna Ivanova, MIT

September 13, 2023
The rise of AI psychology

Theory of Mind May Have Spontaneously Emerged in Large Language Models
Michal Kosinski

Assessing Working Memory Capacity of ChatGPT
Dongyu Gong* (dongyu.gong@psy.ox.ac.uk)

Identifying and Manipulating the Personality Traits of Language Models
Graham Caron and Shashank Srivastava

AI Psychometrics: Using psychometric inventories to obtain psychological profiles of large language models
Max Pellert, Clemens M. Lechner, Claudia Wagner, Beatrice Rammstedt & Markus Strohmaier

Communication
The Political Biases of ChatGPT
David Rozado

Can AI language models replace human participants?
Danica Dillion,1 Niket Tandon,2 Yuling Gu,2 and Kurt Gray1,*,@
The Turing test

My name is Tom

My name is Tom
When evaluating LLM capabilities, we should dissociate language and thought.
Roadmap

- Introduction
- Formal vs. functional linguistic competence
- Formal competence: grammar
- Functional competence
  - math
  - world knowledge
- Case study: event knowledge
- Toward better models
Roadmap

- Introduction
- Formal vs. functional linguistic competence
  - Formal competence: grammar
  - Functional competence
    - math
    - world knowledge
- Case study: event knowledge
- Toward better models
Large Language Models

Large language models today are...

- neural networks
- trained on large amounts of text

on the word-in-context prediction task

The fox chased the XXXX

The fox XXXX the rabbit.

and sometimes fine-tuned on additional tasks
Large Language Models

These models are becoming very good at generating paragraphs of text in response to a prompt.

Example: ChatGPT (released in December 2022)

AN

Explain what language models are to a five-year-old
Large Language Models

These models are becoming very good at generating paragraphs of text in response to a prompt.

Exciting features:

• Novel sentences (not taken verbatim from the web)
• Grammatically correct
• (Seemingly) meaningful

Explain what language models are to a five-year-old

A language model is a way for a computer to understand and generate human language. It's like a program that can read and write words and sentences, and knows what sounds right and what doesn't. This can help the computer do things like answer questions, translate between languages, or even write stories.
Roadmap

• Introduction

• Formal vs. functional linguistic competence
  • Formal competence: grammar
  • Functional competence
    • math
    • world knowledge

• Case study: event knowledge

• Toward better models
Language and the brain

Language processing in the brain takes place within a separate network.

Words, phrases, sentences

Listening and reading

Speaking and writing

Fedorenko et al, 2010, 2011; Scott et al, 2017; Hu, Small et al, 2022; etc, etc
Language and the brain

Language areas show little/no response when we engage in diverse thought-related activities.

Response in the language areas

- Logical reasoning
- Conceptual knowledge
- Math
- Problem solving
- Physical reasoning
- Social reasoning
- etc.

2+17 ⇒

slide adapted from Ev Fedorenko; for a review, see Fedorenko & Varley, 2016
Language areas can be **damaged** with little/no effect on thought-related activities.

*Sample patients’ lesions:*

*slide adapted from Ev Fedorenko; for a review, see Fedorenko & Varley, 2016*
Formal and functional linguistic competence

- Core language knowledge
- Semantic tasks
- Social knowledge
- General cognitive tasks
- Situation modeling
- World knowledge

Mahowald, Ivanova et al, arXiv
Formal and functional linguistic competence

- Core language knowledge
- Semantic tasks
- Social knowledge
- General cognitive tasks
- Situation modeling
- World knowledge

Mahowald, Ivanova et al, arXiv
Formal and functional linguistic competence

**FORMAL COMPETENCE**
(language-specific)

- core language knowledge

**FUNCTIONAL COMPETENCE**
(non-language-specific)

- semantic tasks
- social knowledge
- general cognitive tasks
- situation modeling
- world knowledge

*Mahowald, Ivanova et al, arXiv*
<table>
<thead>
<tr>
<th>FORMAL COMPETENCE</th>
<th>FUNCTIONAL COMPETENCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>(language-specific)</td>
<td>(non-language-specific)</td>
</tr>
<tr>
<td>The keys to the cabinet <strong>are</strong> on the table.</td>
<td>Six birds were sitting on a tree. Three flew away, but then one came back. There are now <strong>four</strong> birds.</td>
</tr>
</tbody>
</table>
Roadmap

• Introduction
• Formal vs. functional linguistic competence
  • **Formal competence: grammar**
  • Functional competence
    • math
    • world knowledge
• Case study: event knowledge
• Toward better models
# Formal competence in large language models

<table>
<thead>
<tr>
<th>Statistical Language Modeling</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon (1940s and 1950s): ngrams, statistical language models, prediction engines</td>
<td>Chomsky (1957): “Despite the undeniable interest and importance of semantic and statistical studies of language, they appear to have no direct relevance to the problem of determining or characterizing the set of grammatical utterances.”</td>
</tr>
<tr>
<td>Rumelhart &amp; McClelland (1986): “Thus the behavior of the model was lawful even though it contained no explicit rules.”</td>
<td>Pinker &amp; Prince (1988): “We conclude that connectionists' claims about the dispensability of rules in explanations in the psychology of language must be rejected”</td>
</tr>
<tr>
<td>2010s and 2020s: Manning et al. (2020): “However, we demonstrate that modern deep contextual language models learn major aspects of [linguistics] structure, without any explicit supervision.”</td>
<td>Everaert et al. (2015) on structures, not strings: “Applying analytical or statistical tools to huge corpora of data in an effort to elucidate the intriguing properties of parasitic gaps will not work.”</td>
</tr>
<tr>
<td>2023: Steve Piantadosi says “Modern language models refute Chomsky’s approach to language”</td>
<td>Roni Katzir (2023): “While LLMs are successful as engineering tools, we saw that they are very poor theories of human linguistic cognition.”</td>
</tr>
</tbody>
</table>

adapted from Kyle Mahowald
Formal competence in large language models

The keys to the cabinet are on the table

Complete the sentence "Far beyond the houses, the person who has the cats that the dogs with the cute faces and fluffy tails chased"

is probably worried about the safety of their beloved feline pets and hoping they can escape unharmed.

adapted from Kyle Mahowald
Formal competence in large language models

Systematic evaluation:
BLiMP (the Benchmark of Linguistic Minimal Pairs for English; Warstadt et al, 2019)

<table>
<thead>
<tr>
<th>Phenomenon</th>
<th>N</th>
<th>Acceptable Example</th>
<th>Unacceptable Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaphor agr.</td>
<td>2</td>
<td>Many girls insulted themselves.</td>
<td>Many girls insulted herself.</td>
</tr>
<tr>
<td>Arg. structure</td>
<td>9</td>
<td>Rose wasn’t disturbing Mark.</td>
<td>Rose wasn’t boasting Mark.</td>
</tr>
<tr>
<td>Binding</td>
<td>7</td>
<td>Carlos said that Lori helped him.</td>
<td>Carlos said that Lori helped himself.</td>
</tr>
<tr>
<td>Control/raising</td>
<td>5</td>
<td>There was bound to be a fish escaping.</td>
<td>There was unable to be a fish escaping.</td>
</tr>
<tr>
<td>Det.-noun agr.</td>
<td>8</td>
<td>Rachelle had bought that chair.</td>
<td>Rachelle had bought that chairs.</td>
</tr>
<tr>
<td>Ellipsis</td>
<td>2</td>
<td>Anne’s doctor cleans one important book and Stacey cleans a few.</td>
<td>Anne’s doctor cleans one book and Stacey cleans a few.</td>
</tr>
<tr>
<td>Filler-gap</td>
<td>7</td>
<td>Brett knew what many waiters find.</td>
<td>Brett knew that many waiters find.</td>
</tr>
<tr>
<td>Irregular forms</td>
<td>2</td>
<td>Aaron broke the unicycle.</td>
<td>Aaron broken the unicycle.</td>
</tr>
<tr>
<td>Island effects</td>
<td>8</td>
<td>Whose hat should Tonya wear?</td>
<td>Whose should Tonya wear hat?</td>
</tr>
<tr>
<td>NPI licensing</td>
<td>7</td>
<td>The truck has clearly tipped over.</td>
<td>The truck has ever tipped over.</td>
</tr>
<tr>
<td>Quantifiers</td>
<td>4</td>
<td>No boy knew fewer than six guys.</td>
<td>No boy knew at most six guys.</td>
</tr>
<tr>
<td>Subject-verb agr.</td>
<td>6</td>
<td>These casserole disgusts Kayla.</td>
<td>These casseroles disgusts Kayla.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPT-2</td>
<td>81.5</td>
</tr>
<tr>
<td>Human</td>
<td>88.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>davinci (175B)</td>
<td>0.84</td>
</tr>
<tr>
<td>GPT-NeoX (20B)</td>
<td>0.839</td>
</tr>
<tr>
<td>TNLG v2 (6.7B)</td>
<td>0.835</td>
</tr>
<tr>
<td>GPT-J (6B)</td>
<td>0.834</td>
</tr>
</tbody>
</table>
Conclusion: large language models have essentially mastered formal linguistic competence.
Roadmap

• Introduction
• Formal vs. functional linguistic competence
• Formal competence: grammar
• **Functional competence**
  • math
    • world knowledge
• Case study: event knowledge
• Toward better models
Six birds were sitting on a tree. Three flew away, but then one came back.

How many birds are there now?
Large language models and math/reasoning

If behavior is identical, why should we care?

It’s not identical - the mechanism affects the system's ability to

(a) generalize

(Example from Chris Cundy)

(b) integrate verbal and nonverbal input

Project Euler

DALL-E

“five elephants playing, a cartoon"
Large language models and math/reasoning

Brown et al (2020)

Dziri et al (2023)
Large language models and math/reasoning

Gary Marcus, 2023:

I went to out to lunch the other day. Sam, Alice, and Barbara were there, plus three other people I didn’t know. How many people were there in total?

There were a total of 8 people at lunch.
Conclusion: large language models diverge substantially from humans in their mathematical reasoning mechanisms
Roadmap

- Introduction
- Formal vs. functional linguistic competence
- Formal competence: grammar
- **Functional competence**
  - math
  - world knowledge
- Case study: event knowledge
- Toward better models
Large language models and world knowledge

More than language:
- embedded in context (place and time are inferred)
- the same effect can be achieved without language input (e.g. by looking out the window)
- preserved in the face of language impairment
Large language models and world knowledge

Huth et al., 2016
Large language models and world knowledge

If behavior is identical, why should we care?

It’s not identical - the mechanism affects the system's ability to

(a) generalize

(b) integrate verbal and nonverbal input

see also Ettinger, 2020; Kassner & Schütze, 2020; Talmor et al, 2020

What is bigger, a microwave or a snowman?

I apologize, but I do not actually have a meaningful way to compare the size of a microwave and a snowman. They are disparate objects with no common basis for size comparison.

"a fox is chasing a rabbit, a cartoon"
Large language models and world knowledge

If behavior is identical, why should we care?

It’s not identical - the mechanism affects the system’s ability to

(c) be consistent

(d) be factually accurate

see also Elazar et al, 2021; Ribeiro et al, 2019; Ravichander et al, 2020, etc

provide some references on the topic "agriculture on Mars"

   • This book explores the challenges and opportunities of cultivating plants in space, including on Mars.

2. "Mars Colonization: Ethical Considerations on Humanity's Next Frontier" by Tony Milgrom.
   • This book discusses the ethical implications of colonizing Mars, including the role of agriculture in sustaining human life.

Large language models and world knowledge

Conclusion: large language models diverge substantially from humans in their world knowledge representations and use
Roadmap

• Introduction

• Formal vs. functional linguistic competence

• Formal competence: grammar

• Functional competence
  • math
  • world knowledge

• Case study: event knowledge

• Toward better models
Generalized Event Knowledge (GEK; McRae & Matsuki 2009)
- storage of **templates** of **common events** observed in the world

*Single words* activate GEK

<table>
<thead>
<tr>
<th>arrest</th>
<th>cop</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>arrest</td>
</tr>
</tbody>
</table>

Words rapidly **combine** to cue specific concepts in GEK

The **journalist** checked the **spelling**.
The **mechanic** checked the **brakes**.

Does GEK rely on language processing in humans?

*sx adapted from Carina Kauf*
Language and event knowledge

SENTENCES

The cop is arresting the criminal.

The criminal is arresting the cop.

PICTURES

TASKS

SEMANTIC
(plausible or implausible?)

PERCEPTUAL
(moving left or right?)

Effect size

Ivanova et al, 2021
Language network & event semantics

The results generalize to other event semantics experiments, but not to object semantics.

Expt. 1  Expt. 2  Expt. 3

Ivanova et al, in prep

Naturalistic event viewing/listening

Sueoka*, Paunov*, Ivanova et al, bioRxiv

Object categorization

Benn*, Ivanova* et al, 2023
Language and event knowledge

- Two participants with global aphasia (PR and SA)
- 12 age-matched controls

**SENTENCES**

The cop is arresting the criminal.

The criminal is arresting the cop.

**PICTURES**

![Image of two individuals, one arresting the other]

**Graph**

- Controls
- PR
- SA

Accuracy

- Picture plausibility judgments
- Sentence picture matching

Ivanova et al, 2021
Two participants with global aphasia (PR and SA)
12 age-matched controls

The language network is recruited but not required for event semantics.
Language models and event knowledge

Does event knowledge naturally arise in language models?

Approach: minimal sentence pairs

<table>
<thead>
<tr>
<th>SENTENCE</th>
<th>The fox chased the rabbit.</th>
<th>The rabbit chased the fox.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCORE</td>
<td>0.8</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Kauf*, Ivanova* et al, arXiv

Kauf*, Ivanova* et al, arXiv
Language models and event knowledge

Animate-Inanimate, impossible
The teacher bought the laptop.
The laptop bought the teacher.

Animate-Animate, unlikely
The fox chased the rabbit.
The rabbit chased the fox.

“the gap between the impossible and the unlikely”

Kauf*, Ivanova* et al, arXiv
Language models and event knowledge

**Animate-Inanimate, impossible**

The teacher bought the laptop.
The laptop bought the teacher.

**Animate-Animate, unlikely**

The fox chased the rabbit.
The rabbit chased the fox.

---

**Selectional restrictions**

= formal competence

**Graded event knowledge**

= functional competence

---

Kauf*, Ivanova* et al, arXiv
Roadmap

- Introduction
- Formal vs. functional linguistic competence
- Formal competence: grammar
- Functional competence
  - math
  - world knowledge
- Case study: event knowledge
- Toward better models
Implications for future models

The formal/functional competence distinction has two implications:

1. **Modular models**
   - architectural modularity
   - emergent modularity
Implications for future models

The formal/functional competence distinction has two implications:

1. **Modular models**
   - architectural modularity
   - emergent modularity

2. **Targeted benchmarks**
   - formal competence: BLiMP, SyntaxGym, etc.
   - functional competence: ?
Summary

• Formal competence = knowledge of linguistic rules and patterns

• Functional competence = non-language-specific skills required for real-life language use

• This distinction (grounded in neuroscience) helps to clarify the discourse around LLMs & to develop targeted assessments of their abilities.
Thanks to…

Kyle Mahowald  Carina Kauf  Ev Fedorenko

and all the other co-authors
Thank you for listening!

Email: annaiv@mit.edu

Twitter: @neuranna