The Symbol Grounding Problem: Conceptual And (A Few) Empirical Aspects

AIAI 2024, University of Göttingen

Dr. Reto Gubelmann, Universities of St. Gallen and Zurich

17 May 2024

Introduction: The Symbol Grounding Problem (SGP)

Introduction: The Symbol Grounding Problem

- The main claim guiding the presentation: **We do not need to worry about the symbol grounding problem regarding LLMs**.
- So, very briefly and preliminarily: *What is the Symbol Grounding Problem? "How is symbol meaning to be grounded in something other than just more meaningless symbols? This is the symbol grounding problem." (Harnad, 1990, p. 340)*
- Compare: Searle in the Chinese Room.
- **Intentionality**, understood as the aboutness or meaningfulness of language, is often taken to require symbol grounding.

Introduction: The Symbol Grounding Problem (SGP)

Bender & Koller's Application of the SGP to LLMs

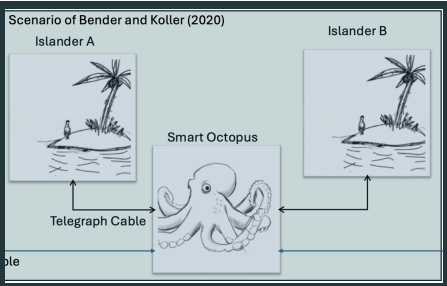
SGP: Outlining Some Philosophical Underpinnings

But, Does *Natural* Language, and hence, do LLMs, Need Symbol Grounding?

Wrapping Up

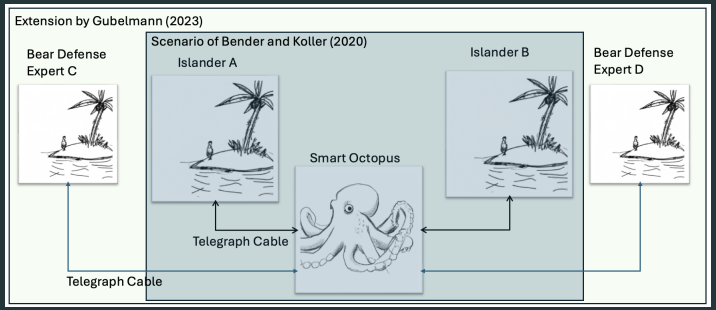
Bender & Koller's Application of the SGP to LLMs

Bender & Koller's Application of the SGP to LLMs



"Finally, A faces an emergency. She is suddenly pursued by an angry bear. She grabs a couple of sticks and frantically asks B to come up with a way to construct a weapon to defend herself. Of course, O has no idea what A "means". Solving a task like this ***requires the ability to map accurately between words and real-world entities*** *[emph RG] (as well as reasoning and creative thinking). It is at this point that O would fail the Turing test, if A hadn't been eaten by the bear before noticing the deception." (Bender and Koller, 2020, p. 5189)*

Responding to Bender & Koller and how this might generalize

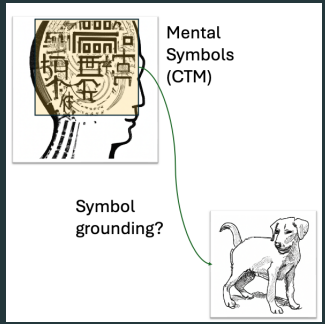


- Same data suffices – no mapping between words and real-world entities required.
- Claim: This generalizes: LLMs don't need any symbols grounded to perform any linguistic task whatsoever.

SGP: Outlining Some Philosophical Underpinnings

Symbolic Computational Theory of the Mind (CTM): The First Habitat of the SGP

- The mind is essentially a computer: Not essentially connected to its wetware, identified by the software that runs on it.
- The mind is a manipulator of mental symbols.
- But what connects **these** (non-natural-language) symbols to their (mind-external) meanings?
- Broadly reminiscent of Cartesianism.



Model-Theoretic Semantics (MTS): The Second Habitat of the SGP (early stages)

- MTS a kind of truth-conditional semantics: To understand a sentence is to know the conditions under which it is true.
- Uses the basic idea of interpreting formal languages to interpret natural language to determine when a sentence is true.
- But how do you map, say, individual constants (e.g., nouns) to the real-world entities that they signify?

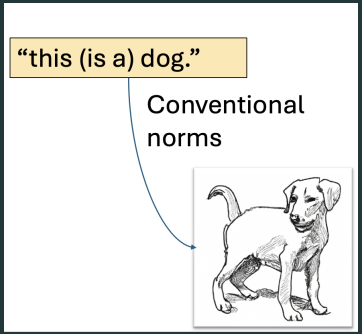
But, Does *Natural* Language, and hence, do LLMs, Need Symbol Grounding?

Charting the Landscape

- Symbolic CTM and MTS both provide a conceptual background from which the SGP makes sense.
- However, one of them is a specific conception of the mind (as opposed to natural language), and the other is an abstraction from natural language.
- From a pragmatic perspective, what MTS abstracts from most clearly (and perhaps most consequentially) is the conventional, norm-governed nature of natural language.

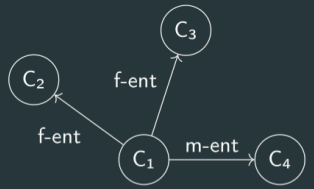
How Expressions of Natural Language Receive their Meaning (and Grounding)

- Expressions in natural language, such as "this is a dog." receive their meaning through norm-governed use: "This is how you should use this expression!"
- It seems that, in contrast to mental symbols at the core of symbolic CTM, or to the abstraction of natural language considered by MTS, in this everyday use, we can easily explain how such symbols are grounded.



Some More Details: A Brandomian-Pragmatist View on Linguistic Meaning

1. Normative-logical relationships between claims are the basic elements: Commitment to certain claims entails and precludes commitment to others.
2. The meaning of a claim is determined by its situation in this inferential network.
3. The meaning of a concept is derived from the inferential difference it makes when occurring in claims.
4. In this sense, individual concepts come last, not first, and they receive their meaning not by connecting them to bits of world, but by their systematic, inference-altering occurrence in claims.



C_1: No cat gets along well with dogs. C_2: My cat Carl won't get along well with our neighbor's dog Rudy. C_3: If something gets along well with a dog, it's not a cat. C_4: No cat can be friends with dogs.

Why the SGP Does Not Apply to LLMs

- We don't need to (and arguably should not) think of LLMs as minds in the symbolic CTM sense.
- As connectionist systems, it's not impossible, but certainly far-fetched to interpret them using MTS.
- LLMs can pick up the regularities and patterns caused by our following of the norms that govern language use.
- There seems therefore no principled limit to the LLMs' performance in any linguistic task; empirical data has (so far) shown little reason to assume that some deficit in LLM-performance can be ascribed to lack of symbol grounding (compare the Octopus earlier).
- (This does not mean that it understands meaning – on a Brandomian-Pragmatist view, it doesn't, because it merely acts in accordance with rules, but doesn't follow them: It cannot decide to violate them by lying, for instance)

Wrapping Up

Summary: A different Conception of Intentionality and Meaning

What I've tried to do:

- Sketch some of the conceptual presuppositions of the symbol grounding problem, especially the CTM and the MTS.
- Suggest that these conceptual presuppositions don't fit very well with LLMs.
- Sketch a positive alternative to CTM and MTS, according to which natural language expressions receive their meaning from the norms that govern their use, with norms regulating inference at the core.
- Suggest that, on that view, there is no principled limit to the linguistic performance of LLMs, while they still won't understand the meaning of natural language expressions.

PS: How to find out who is right?

- We are discussing here what Carnap (1950) has called framework questions: Hence, they cannot be decided empirically because they co-determine what counts as empirical decision.
- Typical evaluation criteria for the evaluation of frameworks are:

  Fruitfulness Does the framework inspire innovative and empirically sound research?

  Fit with Data Does the picture that the framework paints of its range of phenomena seem to fit the relevant empirical data?
- With regard to fruitfulness, the Brandomian framework still has to prove its value; regarding the fit dimension, it seems advantageous that it, unlike the frameworks where the SGP lives, does not set any limits regarding the performance of LLMs trained "purely on form/language", etc.

PPS: Metaphilosophical Afterthoughts

The paper tries to cash out the metaphors of Bender and Koller (2020), according to which conceptual reflection is a top-down perspective, while empirical research is a bottom-up effort (which might climb up the wrong hill!). It did so by implicitly following impure conceptual analysis, that tries to steer a third way between apriorism and scientism.

Apriorism Empirical science has nothing whatsoever to contribute to philosophical-conceptual questions (Bennett and Hacker, 2003)

Scientism Conceptual analysis is not a distinctive source of insight (Quine, 1980 [1951]).

Impure CA Science can show conceptual analyses are beside the point, CA can help science by clarifying concepts and suggesting alternatives. Compare Glock (2013) and Burge (2010).

References

- Bender, Emily and Alexander Koller (2020). **"Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data".** In: *Proceedings ofthe 58th Annual Meeting ofthe Association for Computational Linguistics*, pp. 5185–5198.
- Bennett, Maxwell and Peter Michael Stephan Hacker (2003). ***Philosophical foundations of neuroscience.*** Vol. 79. Blackwell Oxford.
- Brandom, Robert (1998). ***Making It Explicit. Reasoning, Representing, and Discursive Commitment.*** Harvard University Press.

References ii

- Brandom, Robert (2010). ***Between Saying and Doing: Towards an Analytic Pragmatism.*** Oxford University Press.
- — (2021). ***Articulating Reasons.*** Harvard University Press.
- Burge, Tyler (2010). ***Origins of Objectivity.*** Oxford: Oxford University Press.
- Carnap, Rudolf (1950). **"Empiricism, Semantics, and Ontology".** In: *Revue Internationale de Philosophie* 4.11, pp. 20–40.
- Glock, Hans-Johann (2013). **"Animal Minds: Philosophical and Scientific Aspects".** In: *A Wittgensteinian Perspective on the Use of Conceptual Analysis in Psychology*. Palgrave MacMillan, pp. 130–152.
- Harnad, Stevan (1990). **"The Symbol Grounding Problem".** In: *Physica D: Nonlinear Phenomena* 42, pp. 335–346.

References iii

Quine, Willard Van Orman (1980 [1951]). **"Two Dogmas of Empiricism".** In: *From a Logical Point of View*. Harvard University Press, pp. 20–46.