The role of language in human and machine intelligence

Contributors: Gary Lupyan (organizer lupyan@wisc.edu)[1]; Martin Zettersten (mzettersten@ucsd.edu)[2], Hunter Gentry (hrgentry@ksu.edu)[3], Anna Ivanova (a.ivanova@gatech.edu)[4], Thomas L. Griffiths (tomg@princeton.edu)[5], Sean Trott (sttrott@ucsd.edu)[2]

1. Department of Psychology. University of Wisconsin-Madison **2.** Department of Cognitive Science, UC San Diego
3. Department of Philosophy, Kansas State University **4.** School of Psychology, Georgia Tech
5. Departments of Psychology and Computer Science, Princeton University

Keywords: language and thought; statistical learning; large language models; world models

Introduction

We use language to communicate our thoughts. But is language merely the expression of thoughts, which are themselves produced by other, nonlinguistic parts of our minds? Or does language play a more transformative role in human cognition, allowing us to have thoughts that we otherwise could (or would) not have? Recent developments in artificial intelligence and cognitive science have reinvigorated this old question. Could language hold the key to the emergence of both artificial intelligence and important aspects of human intelligence? The four contributions in this symposium address this question by drawing on behavioral and neural evidence from people, and the remarkable recent developments in AI which appear to show that artificial neural networks trained on language come to have an astonishing range of abilities. Despite the diversity of the speakers' perspectives, the four contributions paint a coherent (if complex) picture: The abilities of large language models (LLMs) serve as an existence proof of what is—in principle—learnable from language, and act as a stress test of cognitive theories. The evidence of neural dissociation between linguistic and conceptual processing points to the multiple realizability of human-like cognition. Finally, there is an acknowledged need for systematic research on how the successes and failures of LLMs inform our understanding of human cognition.

Delineating concepts and language in brains and in machines

Anna Ivanova

Much of what humans know about the world they learn through language. Can we equate factual and/or distributional information extracted from language with conceptual knowledge? And does the tight link between language and concepts mean that the brain has shared computational circuits for linguistic and conceptual processing?

To answer the first question, we can turn to large language models (LLMs). LLMs serve as a valuable testbed for establishing what conceptual knowledge is easily learnable through statistical learning over large language corpora. I will show that distributional language knowledge allows LLMs to easily distinguish possible and impossible event schemas, but their knowledge of likely vs unlikely events is less robust. I will then present a framework for systematically assessing LLMs knowledge of specific concepts, sourced from domains of knowledge that have long been studied by cognitive scientists. This systematic evaluation shows that LLMs consistently show knowledge deficits in physical and spatial knowledge domains, indicating that language alone might be insufficient for acquiring and/or effectively using knowledge about the physical world (although it contains a surprising amount of information about social concepts).

To answer the second question, we can leverage cognitive neuroscience. I will discuss fMRI evidence that indicates the existence of a network of brain regions that respond to semantic tasks, performed on either linguistic or pictorial stimuli. These semantic demand regions are adjacent to, but not identical with, the language-responsive regions in the brain and are not equivalent to the domain-general task-responsive (multiple demand) brain network. This line of work suggests that language and semantic processing rely on interconnected but separate neural circuits.

Finally, I will conclude by reviewing the language-concepts relationship through the framework of formal vs. functional linguistic competence (Mahowald et al., 2024). A distinction between formal competence—knowledge of linguistic rules and patterns—and functional competence—understanding and using language in the world—can help clarify the relationship between language processing, as performed by the language network, and conceptual processing, as performed by semantic demand regions (among others). Thus, despite the necessity of conceptual knowledge for successful language use, language and concepts likely rely on separate cognitive mechanisms.

Exploring the limits of language with large language models

Thomas L. Griffiths

The remarkable success of large language models as a basis for creating artificial intelligence systems is arguably a demonstration that language might play an even more important role in intelligence than we might have expected. However, these models also provide an opportunity to demonstrate the limits of language. I will talk about two sets of results that combine large language models with the methods of cognitive science to explore these limits. First, as systems that have had no experience of the world beyond the linguistic descriptions they are trained upon, large language

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

models provide an opportunity to evaluate how much of sensory experience might be captured from language alone (Marjieh et al., 2024). Second, we can use cognitive science to identify cases where asking large language models to produce additional text — a "chain of thought" — results in decreased performance (Liu et al., 2024). Each case highlights a way in which language falls short of capturing all of human cognition, and hence a potential lacuna in the abilities of AI systems based on large language models.

Do we know enough to know what language models know?

Sean Trott

Isolating the causal role of linguistic input in shaping human cognition has historically been extremely difficult. Recent advances in large language models (LLMs) have made it more tractable to test a narrow version of this question: what can you learn from language alone? With appropriate care, LLMs could perhaps be used to provide evidence about whether exposure to language is sufficient to produce a range of cognitively interesting behaviors.

I begin by examining whether a system trained on the statistics of language produces behavior consistent with the ability to reason about false beliefs, one facet of Theory of Mind (Trott et al., 2023). We find mixed results: language does confer *some* sensitivity to the belief states of characters in a story; but this sensitivity falls short of most humans tested. The picture is further complicated by interpretive challenges, i.e., whether an LLM's performance on a task designed for humans ought to be taken as evidence for the construct that the task ostensibly indexes.

I then turn to more general *epistemological challenges* relating to the use of LLMs as "model organisms" to inform debates about human cognition. Researcher intuitions vary as to whether the same task indexes the same thing for humans and LLMs or whether the task exhibits "differential construct validity". It is also unclear what constitutes generalization and which are "mere" pattern-matching. Finally, unlike the study of non-human animals, LLMs do not share biological continuity with humans and we therefore lack strong theoretical principles to guide decisions about which LLMs can be used as "models" of human cognition for which tasks. I conclude by suggesting that these challenges are crucial opportunities for methodological and theoretical refinement: given the increasing ubiquity of LLMs in cognitive science research, it is essential that we develop the conceptual and analytical toolkit to figure out what they can and can't do.

How important is language for human-like intelligence?

Gary Lupyan, Hunter Gentry, Martin Zettersten

Notwithstanding the hype of impending artificial general intelligence, training neural networks on large amounts of natural language has been a far more successful endeavor than anyone imagined. What does it mean that exposure to language seems to endow general purpose neural networks with such a wide range of skills—pragmatics, theory of mind, categorization, some forms of reasoning? We argue that there are two main lessons. First, self-supervised prediction is far more powerful for learning structured representations than previously supposed (Arcas, 2022; Lupyan & Clark, 2015). Second, it is not a coincidence that it was training on natural language that led to these breakthroughs. Rather, predicting language acts as a powerful guiding and constraining force on general-purpose cognitive mechanisms resulting in the learning of abstract and generative 'cognitive chunks' that promote inference and reasoning (Lupyan & Zettersten, 2021). The uniquely open-ended nature of language means that it can be used to convey everything from how we feel, to recipes, to scientific findings. Reducing prediction error across these varied domains turns out to be an effective strategy for gaining a wide range of expertise.

Adopting this perspective makes the successes of artificial neural networks trained on language less surprising: Despite the vast differences between LLMs and human minds, language appears to help both.

References

Arcas, B. A. y. (2022, February 16). Do large language models understand us? *Medium*. https://medium.com/@blaisea/do-large-language-models-understand-us-6f881d6d8e75

Liu, R., Geng, J., Wu, A. J., Sucholutsky, I., Lombrozo, T., & Griffiths, T. L. (2024). *Mind Your Step (by Step): Chain-of-Thought can Reduce Performance on Tasks where Thinking Makes Humans Worse* (No. arXiv:2410.21333).

Lupyan, G., & Clark, A. (2015). Words and the World: Predictive coding and the language-perception-cognition interface. *Current Directions in Psychological Science*, *24*(4), 279–284. https://doi.org/10.1177/0963721415570732

Lupyan, G., & Zettersten, M. (2021). Does Vocabulary Help Structure the Mind? In M. D. Sera & M. A. Koenig (Eds.), *Minnesota Symposia on Child Psychology* (pp. 160–199). https://doi.org/10.1002/9781119684527.ch6

Mahowald, K., Ivanova, A. A., Blank, I. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2024). Dissociating language and thought in large language models. *Trends in Cognitive Sciences*, *28*(6), 517–540. https://doi.org/10.1016/j.tics.2024.01.011

Marjieh, R., Sucholutsky, I., van Rijn, P., Jacoby, N., & Griffiths, T. L. (2024). Large language models predict human sensory judgments across six modalities. *Scientific Reports*, *14*(1), 21445. https://doi.org/10.1038/s41598-024-72071-1

Trott, S., Jones, C., Chang, T., Michaelov, J., & Bergen, B. (2023). Do Large Language Models Know What Humans Know? *Cognitive Science*, *47*(7), e13309. https://doi.org/10.1111/cogs.13309