Can Natural Language Models be Part of Linguistic Communities?

Research Proposal

Alexander Niederklapfer, a.niederklapfer@lse.ac.uk

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Background

Current Natural Language Models (NLMs) like GPT-4, BERT, and others show impressive performance on various linguistic tasks. [1]

But do they "understand" the "meaning" of texts?

Many conceptions of meaning seem to be entangled with a notion of <u>linguistic community</u>:

- Semantic externalism: words refer through causal history in a linguistic community [2,3]

- Putnam: words get meaning/reference through deferral to experts in a linguistic community [4,5]

- Wittgenstein: meaning determined by "use" –arguably use within a linguistic community ([1] employs this notion, without explicitly acknowledging the dependence on linguistic communities)

Thus, one way to address the problem of understanding in NLMs is asking the question whether they can be part of linguistic communities.

Current arguments in favour	Current arguments against
Evidence that NLMs can invent new words [6]	NLMs lack intentions [4]
No in-principle reasons against [2]	NLMs lack long-term web of concepts [4]

Possible research directions

- Does fine-tuning provide the grounds for including NLMs in a language community?

- What is a language community, after all? What do linguists have to say about this?

- Does being part of a language community provide the *grounding* of meaning that some think is necessary? [7]

- If they are not active participants, we could think that NLMs are "cultural technologies", like photocopiers. ([6] call this "Bibliotechnism")
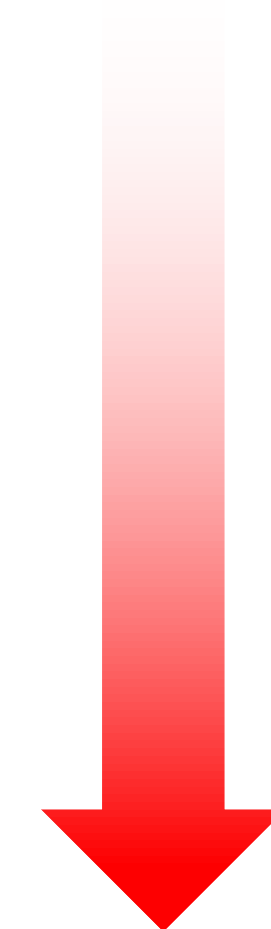
An analogy

Someone learns Ancient Greek through reading texts by long dead authors.

They start producing texts that are passively consumed by other scholars.

Other scholars actively engage with them in Ancient Greek. They add new nuances to words, maybe invent new words.

Not part

Part of a linguistic community?

[1] Gubelmann, Reto. 2023. A Loosely Wittgensteinian Conception of the Linguistic Understanding of Large Language Models like BERT, GPT-3 and ChatGPT. *Grazer Philosophische Studien* 99.4, 482-523.

[2] Mandelkern M. and Tal Linzen. 2024. Do Language Models' Words Refer? *Computational Linguistics*.

[3] Kripke, Saul. 1980. *Naming and Necessity*. Harvard University Press, Cambridge, MA.

[4] Millière, Raphael and Cameron Bruckner. 2024. A Philosophical Introduction to Language Models – Part I: Continuity with Classic Debates. arXiv:2401.03910

[5] Putnam, Hilary. 1981. *Reason, Truth, and History*. Cambridge University Press.

[6] Ledermann, Harvey and Kyle Mahowald. 2024. Are language models more like libraries or like librarians? Bibliotechnism, the novel reference problem and the attitudes of LLMs. arXiv:2401.04854.

[7] Mollo, Dimitri C. and Raphael Millière. 2023. The Vector Grounding Problem. arXiv:2304.01481