Tel Aviv University
School of Philosophy, Linguistics and Science Studies,
Department of Linguistics

THURSDAY INTERDISCIPLINARY COLLOQUIUM

Thursday 01/01/2026
16:15-17:45
Webb 103

Imry Ziv, Tel Aviv University

Biasless Language Models Learn Unnaturally: How Large Language Models Fail to Distinguish the Possible from the Impossible

Are large language models (LLMs) sensitive to the distinction between humanly possible languages and humanly impossible languages? This question is taken by many to bear on whether LLMs and humans share the same innate learning biases. Previous work has attempted to answer it in the positive by comparing LLM learning curves on existing language datasets and on "impossible" datasets derived from them via various perturbation functions. Using the same methodology, we examine this claim on a wider set of languages and impossible perturbations. We find that in most cases, GPT-2 learns each language and its impossible counterpart equally easily, in contrast to previous claims. We also apply a more lenient condition by testing whether GPT-2 provides any kind of separation between the whole set of natural languages and the whole set of impossible languages. By considering cross-linguistic variance in various metrics computed on the perplexity curves, we show that GPT-2 provides no systematic separation between the possible and the impossible. Taken together, these perspectives show that LLMs do not share the human innate biases that shape linguistic typology.

Click here to see the colloquium program.