Large Language Models for Linguistics: Applications and Implications

Convenors: Natalia Levshina natalia.levshina@ru.nl and Nicole Katzir
nicole.katzir@ru.nl (Radboud University)

Keywords: large language models, linguistic theory, data annotation, anthropomorphic language, responsible AI

Workshop description

Large Language Models (LLMs) are models with billions of parameters, trained on vast amounts of text data to learn statistical patterns in language, and able to generate, process, and predict human(-like) text. As discussions at the recent SLE meeting and other venues demonstrate, the rise of LLMs has major consequences for our field. The apparent success of LLMs in producing output that can be difficult to distinguish from human language, as well as their performance on different linguistic tasks, has sparked intense debate about what, if anything, can be inferred for linguistic theory. Others have focused on the potential of LLMs as tools for data annotation or on describing LLM-generated texts as a special "lect". Alongside these scientific debates and studies, the issue of the ethical implications of using LLMs in research remains unresolved.

The goal of this workshop is to bring together linguists, cognitive scientists, computational scientists and other experts to discuss how LLMs intersect with linguistics. Below are the central questions that need to be addressed.

● **What are the repercussions of LLMs for linguistic theory?** The impressive linguistic performance of LLMs has led some scholars to use it as an argument against Chomskyan generative grammar (e.g., Piantadosi 2024) and in favour of usage-based connectionist models (Goldberg 2023), whereas others (Chomsky et al. 2023; Kodner et al. 2023) have claimed that the fact that LLMs can approximate human language does not tell us anything valuable about human language itself. At the same time, it is difficult to deny the fact that LLMs are able to "acquire" nontrivial syntactic generalizations, which cannot be explained by simple heuristics or co-occurrence patterns in the input data (Futrell & Mahowald 2025), such as filler-gap dependencies (Suijkerbuijk et al. 2023) and recursive embedding (Futrell et al. 2019). This raises the question: what are the consequences of these successes for our understanding of how human language is acquired, represented, and processed (cf. Contreras Kallens et al. 2023)?

● **Under what conditions can LLMs be used reliably and ethically in linguistic research?** The potential of some language models has been explored in psycholinguistics (e.g., Wilcox et al. 2023), pragmatics and discourse studies (Chen et al. 2024; Yu et al. 2024), syntax (Ambridge &

Blything 2024; Dunn & Eida 2025), corpus-based language comparison (Koplenig et al. 2025), semantic and grammatical annotation (Levshina et al. 2024; Fuoli et al. 2025; Morin & Larsson 2025), and other subfields, but a more systematic and critical discussion of such uses is needed. In particular, the status of large generative models is under debate. Some researchers dismiss Generative AI entirely as ethically unacceptable due to copyright violations, reproducing and amplifying social biases, environmental impact, exploitation, and other valid concerns (cf. Guest et al. 2025). Others seem to underestimate the extent of the problematic issues surrounding generative models, approaching them as merely another tool. We want to take a more nuanced position and ask if we can employ at least some types of Generative AI in linguistic research in a reliable and ethically acceptable way, for example, as annotation tools or sources of data. Are some models more ethical than others, e.g., open-weight and research-oriented vs. closed-weight and proprietary, small and specialized vs. large and general-purpose, and so on? Additionally, can current standards of transparency and reproducibility be upheld in linguistic research based on Generative AI and LLMs?

● **How should we speak about LLMs?** The linguistic framing of AI – for example, as a tool or a companion – guides social attitudes and behaviours towards these technologies (Petricini 2025). One often hears that LLMs “understand”, “learn”, “think”, “reason”, or “hallucinate”. Not only is such anthropomorphic language erroneous, but it can also lead to exploitation of users’ emotional dependence on AI, misplaced trust, decreasing accountability of Big Tech, and other negative consequences (DeVrio et al. 2025; Placani 2024). It falls to linguists to analyze and challenge such language use, especially in scientific communication.

● **What are the distinctive features of LLM output?** The state-of-the-art LLMs have essentially passed the Turing test, being indistinguishable from human language in different settings, such as textual conversations (Jones & Bergen 2025) and essay writing (Herbold et al. 2023). In fact, they can sound more human than humans as a result of exploiting users’ flawed heuristics about human language (Jakesch et al. 2022). However, some corpus-based studies have managed to identify “fingerprints” of several models (Dentella et al. 2025), especially instruction-tuned ones (Reinhart et al. 2024). Identifying such statistical patterns is crucial both for understanding how these models differ from human language and for developing reliable methods of distinguishing LLM-generated text from human-produced writing.

● **How to solve the data bottleneck?** LLMs require huge amounts of training data, which is only available for relatively few major languages. As a result, while LLMs tend to excel in English, their performance in low-resource languages struggles (Li et al. 2024; Rahman et al. 2024). The same applies to non-standard linguistic varieties, such as dialects and sociolects (Smith et al. 2025). Consequently, the resources available for the speakers of these varieties, as well as for researchers working on them, are limited. Thus, it has been argued that LLMs reflect standard language ideology, which posits hierarchies according to which some language varieties are "better" and "correct" (Smith et al. 2025). How to address and solve this language representation bias?

References

Ambridge, Ben & Lewis Blything. 2024. Large language models are better than theoretical linguists at theoretical linguistics. *Theoretical Linguistics* 50(1–2). 33–48.

Chen, Xi, Jun Li & Yuting Ye. 2024. A feasibility study for the application of AI-generated conversations in pragmatic analysis. *Journal of Pragmatics* 223. 14–30. https://doi.org/10.1016/j.pragma.2024.01.003.

Chomsky, Noam, Ian Roberts & Jeffrey Watumull. 2023. Noam Chomsky: The false promise of ChatGPT. The New York Times.

Contreras Kallens, Pablo, Ross Deans Kristensen-McLachlan & Morten H. Christiansen. 2023. Large Language Models demonstrate the potential of statistical learning in language. *Cognitive Science* 47(3). e13256. https://doi.org/10.1111/cogs.13256.

Dentella, Vittoria, Weihang Huang, Silvia Angela Mansi, Jack Grieve & Evelina Leivada. 2025. ChatGPT-generated texts show authorship traits that identify them as non-human. arXiv. https://doi.org/10.48550/ARXIV.2508.16385.

DeVrio, Alicia, Myra Cheng, Lisa Egede, Alexandra Olteanu & Su Lin Blodgett. 2025. A taxonomy of linguistic expressions that contribute to anthropomorphism of language technologies. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–18. Yokohama Japan: ACM. https://doi.org/10.1145/3706598.3714038.

Dunn, Jonathan & Mai Mohamed Eida. 2025. LLMs Learn Constructions That Humans Do Not Know. arXiv:2508.16837. https://doi.org/10.48550/arXiv.2508.16837.

Fuoli, Matteo, Weihang Huang, Jeannette Littlemore, Sarah Turner & Ellen Wilding. 2025. Metaphor identification using large language models: A comparison of RAG, prompt engineering, and fine-tuning. *arXiv*. https://doi.org/10.48550/arXiv.2509.24866

Futrell, Richard, Ethan Wilcox, Takashi Morita, Peng Qian, Miguel Ballesteros & Roger Levy. 2019. Neural language models as psycholinguistic subjects: Representations of syntactic state. In *Proceedings of the 2019 Conference of the North*, 32–42. Minneapolis, Minnesota: Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1004

Futrell, Richard & Kyle Mahowald. 2025. How Linguistics Learned to Stop Worrying and Love the Language Models. arXiv:2501.17047. https://doi.org/10.48550/arXiv.2501.17047

Goldberg, Adele E. 2024. Usage-based constructionist approaches and large language models. *Constructions and Frames* 16(2). 220–254. https://doi.org/10.1075/cf.23017.gol

Guest, Olivia, Marcela Suarez, Barbara Müller, Edwin van Meerkerk, Arnoud Oude Groote Beverborg, Ronald de Haan, Andrea Reyes Elizondo, et al. 2025. Against the uncritical adoption of ‘AI’ technologies in academia. Zenodo. https://doi.org/10.5281/zenodo.17065099.

Herbold, Steffen, Annette Hautli-Janisz, Ute Heuer, Zlata Kikteva & Alexander Trautsch. 2023. AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays. arXiv. https://doi.org/10.48550/ARXIV.2304.14276

Jakesch, Maurice, Jeffrey T. Hancock & Mor Naaman. 2022. Human heuristics for AI-generated language are flawed. *PNAS* 120(11). e2208839120. https://doi.org/10.1073/pnas.2208839120

Jones, Cameron R. & Benjamin K. Bergen. 2025. Large Language Models pass the Turing test. arXiv. https://doi.org/10.48550/ARXIV.2503.23674

Kodner, Jordan, Sarah Payne & Jeffrey Heinz. 2023. Why linguistics will thrive in the 21st century: A reply to piantadosi (2023). arXiv. https://doi.org/10.48550/ARXIV.2308.03228

Koplenig, Alexander, Sascha Wolfer, Jan Oliver Rüdiger & Peter Meyer. 2025. Human languages trade off complexity against efficiency. *PLOS Complex Systems* 2(2). e0000032. https://doi.org/10.1371/journal.pcsy.0000032

Levshina, Natalia, Maria Koptjevskaja-Tamm Maria & Robert Östling. 2024. Revered and reviled: a sentiment analysis of female and male referents in three languages. *Frontiers in Communication* 9. 1266407. https://doi.org/10.3389/fcomm.2024.1266407

Li, Zihao, Yucheng Shi, Zirui Liu, Fan Yang, Ali Payani, Ninghao Liu & Mengnan Du. 2024. Language Ranker: A metric for quantifying LLM performance across high and low-resource languages. arXiv. https://doi.org/10.48550/ARXIV.2404.11553

Morin, Cameron & Matti Marttinen Larsson. 2025. *Large corpora and large language models: a replicable method for automating grammatical annotation. Linguistics Vanguard*, aop. https://doi.org/10.1515/lingvan-2024-0228

Petricini, Tiffany. 2025. The power of language: framing AI as an assistant, collaborator, or transformative force in cultural discourse. *AI & Society*. https://doi.org/10.1007/s00146-025-02586-2

Placani, Adriana. 2024. Anthropomorphism in AI: hype and fallacy. *AI and Ethics 4(3).* 691–698. https://doi.org/10.1007/s43681-024-00419-4

Piantadosi, Steven T. 2024. Modern language models refute Chomsky’s approach to language. In Edward Gibson & Moshe Poliak (eds.), From fieldwork to linguistic theory: A tribute to Dan Everett, 353–414. Berlin: Language Science Press.

Rahman, Abrar, Garry Bowlin, Binit Mohanty & Sean McGunigal. 2024. Towards linguistically-aware and language-independent tokenization for Large Language Models (LLMs). arXiv. https://doi.org/10.48550/ARXIV.2410.03568

Ranjan, Rajesh, Shailja Gupta & Surya Narayan Singh. 2024. A comprehensive survey of bias in LLMs: Current landscape and future directions. arXiv. https://doi.org/10.48550/ARXIV.2409.16430

Reinhart, Alex, Ben Markey, Michael Laudenbach, Kachatad Pantusen, Ronald Yurko, Gordon Weinberg & David West Brown. 2025. Do LLMs write like humans? Variation in grammatical and rhetorical styles. *PNAS* 122 8). e2422455122. https://doi.org/10.1073/pnas.2422455122

Smith, Genevieve, Eve Fleisig, Madeline Bossi, Ishita Rustagi & Xavier Yin. 2024. Standard language ideology in AI-generated language. arXiv. https://doi.org/10.48550/ARXIV.2406.08726

Suijkerbuijk, Michelle, Peter de Swart & Stefan L. Frank. 2023. The learnability of the wh-island constraint in Dutch by a long short-term memory network. *Society for Computation in Linguistics* 6(1). 321-331. https://doi.org/10.7275/D5QF-2H13

Wilcox, Ethan G., Tiago Pimentel, Clara Meister, Ryan Cotterell & Roger P. Levy. 2023. Testing the predictions of surprisal theory in 11 languages. *Transactions of the Association for Computational Linguistics* 11. 1451–1470. https://doi.org/10.1162/tacl_a_00612

Yu, Danni, Luyang Li, Hang Su & Matteo Fuoli. 2024. Assessing the potential of LLM-assisted annotation for corpus-based pragmatics and discourse analysis: The case of apology. *International Journal of Corpus Linguistics* 29(4). 534–561. https://doi.org/10.1075/ijcl.23087.yu