_____

РЕЦЕНЗІЇ

Соломія БУК, Ольга МАКСИМІВ
Львів, Львівський національний університет імені Івана Франка
УДК 81'32/'33:004.8/.9]:001.891

RECONFIGURING LINGUISTIC INQUIRY IN THE AGE OF GENERATIVE AI

Review of the Course **"Large Language Models for Linguists"** by Dr. Natalia Cheilytko, Friedrich Schiller University Jena, supported by German Academic Exchange Service DAAD.

In recent years, the emergence of large language models (LLMs) has brought about a big shift in scientific research generally and in the way how human language is studied in particular. As an academic response to this shift, new relevant disciplines have emerged in foreign and Ukrainian universities, one of which is the course 'Large Language Models for Linguists' at Jena University. Its target audience is Ukrainian teachers and students. Combining computational linguistics and classical philology, the course introduces participants to new methodologies for investigating language variation, semantic change and grammatical structures, making use of generative artificial intelligence tools such as OpenAI's GPT, and it is important in the context of higher education in Ukraine.

Spread over 16 weeks and worth 4 ECTS credits, this course combines theoretical discussion with practical work. Students explored a variety of subjects, ranging from the fundamentals of generative artificial intelligence and tokenisation principles to more advanced applications, including corpus research, lexicography, and discourse analysis.

In addition to modules on translation studies and language teaching, the course surprisingly touches on underexplored domains — like visual data processing and oral language annotation — areas often overlooked in mainstream computational linguistics.

From a pedagogical perspective, the course strikes a good balance between teaching the basics and encouraging student-led research. The final research project encourages students to apply their newly acquired knowledge to authentic linguistic research. The selection of key literature, including seminal works by Jason Way et al. (2022) and Samuel Bowman (2023), demonstrates a close connection to new research in artificial intelligence and natural language processing (NLP). The teaching format was a

_____

_____

combination of synchronous and asynchronous online learning together with consultation tools. It indicates a practical, user-centred design adapted to the needs of Russian-Ukrainian war circumstences.

However, beyond the technical prowess lie embedded reflections on philosophical and ethical issues. What impressed us was the willingness to foreground model fallibility, as well as interpretive bias and the risks of overreliance on generative AI — a rare move in a landscape still enamoured with technological utopianism. In an era where algorithmic mediation is becoming increasingly prevalent, it is imperative to teach students to critically evaluate the outcomes of a LLM application.

From a philosophical perspective, the course invites reflection on the epistemic implications of linguistic modelling. What does it mean to 'understand' language when a machine can imitate meaning without consciousness? How should we interpret the apparent emergence of 'semantic intuition' in transformer-based architectures? While these questions are not explicitly answered in the curriculum, they are raised by its design and scope, thereby fostering intellectual curiosity.

From an applied linguistics perspective, this course provides a new generation of scholars with the tools to reconsider traditional linguistic methodologies. LLMs can assist with annotation, hypothesis generation, pattern recognition and corpus creation. For instance, semantic change detection, which previously required decades of manually curated texts, can now be achieved using scalable embeddings and diachronic probing methods.

However, the course also cautions against the uncritical implementation of LLMs. Students learn to identify 'hallucinations', recognise the biases of models based on training corpora and calibrate results against human experience. The tension between automated processing and human interpretation was an ongoing issue throughout the course, with the two often conflicting — particularly evident in tasks involving corpus annotation and GPT-assisted translation.

The range of topics offered reflects ambitious engagement in linguistic research through LLMs. Studies such as thema-rhema structuring in the media coverage of the Coronavirus pandemic through the lens of SFA, or the application of BERTopic in semantic alignment, demonstrate methodological innovation. Others apply ChatGPT to tasks such as automatic punctuation correction and general test creation using artificial intelligence.

The resulting scholarly work comprised reports from 27 participants and spanned disciplines such as discourse analysis, language teaching, translation, grammar, lexicography and image recognition. Projects focusing on cross-linguistic stylistic adaptation, dialect translation or the interpretation of place names demonstrate promising cross-cultural sensitivity. Attempts to work with historical corpora (e.g. Ukrainian texts from the 16th-17th centuries) or to digitise 20th-century newspapers using optical character recognition highlight a remarkable archival focus. Research on lexical simplification, culinary multilingual mapping and creating frequency dictionaries using

LLMs borders on the experimental. The project on editing poetic translation raises acute theoretical and ethical questions about aesthetic activity — a field that remains largely understudied in contemporary AI discourse. Conversely, studies on phraseology in second language contexts or discursive representations of Ukrainian women during the war demonstrate significant potential for socio-pragmatic understanding. Almost every presentation concludes with the need for careful philological validation of AI results. Ultimately, the diversity of topics confirms the viability of this field, but success depends on careful linguistic theorising that goes beyond technological novelty, as well as careful reflexivity about the epistemic limits of LLMs.

The field of AI is developing very rapidly, so perhaps the connections made during the course will have a synergistic effect. One example of ongoing cooperation is the idea of publishing research results in the peer-reviewed journal LANGUAGE: Codification, Competence, Communication'.

The course design stands out for its integration of diachronic corpus methods into LLM-driven workflows – a combination rarely seen in similar programmes. It balances methodological knowledge with hands-on experimentation, while occasionally prompting ethical reflection.

While the course introduces critical engagement with language models, the depth of theoretical grounding students achieve remains to be seen and warrants follow-up investigation.

What might once have seemed a theoretical convergence between linguistic methodology and computational modeling here takes on a tangible form: the course enables participants not just to observe language through digital tools, but to interrogate its structures with new epistemic sensitivity.

Соломія Бук – професор катедри загального мовознавства Львівського національного університету імені Івана Франка
ORCID: https://orcid.org/0000-0001-8026-3289
Email: solomija@gmail.com

Ольга Максимів – доцент, завідувач катедри сходознавства імені професора Ярослава Дашкевича Львівського національного університету імені Івана Франка
Email: olha.maksymiv@lnu.edu.ua