

Corpus Linguistics & AI Programme

07 May 2026	
Until 12:00	Arrival
12:30	Lunch
13:30	Conference Opening
	Section 1: The study of content and meaning
14:00	Vítor Míguez: Large language models as first-pass filters: from small-sample validation to full-dataset implementation
14:40	Jennifer Ecker, Marc Kupietz, Roman Schneider, Ngoc Duyen Tanja Tu: Thematic content indexing of very large corpora with language models: towards openness, interoperability, and reproducibility
15:20	M. L. Yislén Barboza Hidalgo: ChatGPT as a thematic role tagger in Spanish
16:00	Coffee break
	Section 2: Theoretical implications for corpus linguistics
16:30	João Pedro Padua: Asking it directly: large language models and corpus linguistics responses to language use questions
17:10	Stefan Th. Gries: Towards prototypes/ordinary meaning in legal interpretation: high-cue validity features from AI and synchronic corpus data
17:50	Tobias Bernaisch: Theoretical and practical approaches to the interface between AI and corpus linguistics
18:30	Dinner
08 May 2026	
	Plenary talk
09:00	Andreas Weilinghoff: Human versus AI? Harmonising human and artificial intelligence for corpus linguistics
	Section 3: The corpus-linguistic research process
10:00	Martin Klotz, Thomas Krause, Anke Lüdeling, Carolin Odebrecht, Malte Dreyer: Large language models and the process of corpus linguistic research
10:40	Coffee break
11:10	Valentin Werner, Sabine Weber: Towards a deep-learning-powered POS-tagger of beginner and intermediate EFL interlanguage

11:50	Volker Emmrich: Treebank conversion using AI: a study on the possibilities of deriving universal dependencies from constituency structures of New High German
12:30	Lunch
	Section 4: Applying AI in Applied and Variationist Linguistics
14:00	Margret Mundorf: Perspectivity in legal texts: a comparative analysis of traditional corpus linguistic software and LLMs
14:40	Xiaoyun Huang, Mo Han: From theory to computational realization: applying LLMs to analyze COVID-19 threat discourse in German parliamentary debates
15:20	Torsten Zesch, Katrin Wisniewski: Automated target hypothesis generation in German learner corpora using LLMs
16:00	Coffee break
	Section 5: ChatGPT for Corpus-Linguistic Research?
16:30	Man Zhang: Large language models in metadiscourse identification
17:10	Ruiming Ma: Linguistic subtlety at scale: AI-assisted annotation of Mandarin grammar
17:50	Emily Herman, Pablo Requena: Sociolinguistic competence in an AI-generated corpus: a study of <i>that</i> complementizer
18:30	Dinner
09 May 2026	
	Section 6: Variationist Studies
09:00	Zhao Wenjie: Generative AI-assisted text sentiment analysis applications
09:40	Yawen Zheng, Yuan Li: A domain-specific LLM for error annotation in German learner writing: construction and task-based evaluation
10:20	Coffee break
10:40	Muhammad Shakir, Elen Le Foll: How small can we go? Assessing the reliability of small open-weight LLMs for corpus annotation
11:20	Alfred Lameli: When no model fits: AI-based translation strategies for historical dialect corpora
12:30	Lunch
13:30	Departure