

Andrea Farina, Zinaida Geylikman, Bastien Liétard, Barbara McGillivray

NLP tools and methods for diachronic lexical semantics

Keywords: Lexical Semantics; Historical Linguistics; Lexicology; Lexicography; Corpus Linguistics; NLP

The workshop aims to gather contributions on Natural Language Processing (NLP) tools and methods for the diachronic study of lexical semantics. The focus is on diachrony and lexical semantics, addressing two questions: (1) does the performance of NLP tools vary depending on the historical period and language considered? (2) when are manual or semi-automatic annotations more reliable than NLP methods?

State of the art. Over the past decades, NLP has transformed linguistic research (Nwagwu, 2022). Lemmatization, part-of-speech tagging, and syntactic parsing can now be performed automatically and at scale with high accuracy, including for diachronic corpora (Piotrowski, 2012; Bary et al., 2017; Vatri & McGillivray, 2020). Corpus-based approaches and distributional semantics models are widely integrated into lexical semantics (Boleda, 2020) and networks such as WordNet (Fellbaum, 2006) provide structured representations of semantic relations. In parallel, Linked Open Data initiatives support interoperable ontologies connecting lexical resources across datasets and languages. Together, these approaches enable large-scale modeling of semantic relations, including polysemy and synonymy. NLP tools and Large Language Models (LLMs) have been increasingly applied to lexical semantics, targeting semantic analysis (e.g., Gulordava & Baroni, 2011; Ghinassi et al. 2024; Farina et al., 2026) and metaphor and semantic change (e.g. Periti & Tahmasebi, 2024; Liétard et al., 2024). However, performance varies depending on historical period, genre, and language under consideration, and tasks requiring fine-grained semantic interpretation still benefit from manual or semi-automatic approaches (Glynn, 2009).

Research questions and objectives. This workshop seeks to investigate whether current NLP methods and tools can support the annotation of lexical semantic features (word senses, metaphors, concrete/abstract meanings, etc.) in diachronic corpora. In particular, it aims to assess under which conditions high-quality annotation can be achieved and how this may vary depending on (i) the theoretical framework adopted; (ii) the research objective; (iii) the language under investigation; (iv) the chronological dimensions. A systematic examination of these variables may clarify both the potential and the current limitations of NLP-based approaches to diachronic lexical semantics.

Submission guidelines. We invite contributions presenting research in diachronic lexical semantics that relies, at least partly, on NLP tools and methods. There are no restrictions on theoretical frameworks, languages, or time periods covered. We particularly welcome contributions with a comparative, evaluative, or methodologically reflective perspective that systematically compare human-assisted annotation with fully automatic approaches.

Abstracts should be submitted no later than **May 17, 2026**, via EasyChair: <https://easychair.org/conferences/?conf=llcd2026>. When submitting, please select **WS4. NLP tools and methods for diachronic lexical semantics** in the subsection 'Other Information and Files'. The language of the workshop is English.

Abstracts must be fully anonymized and should not exceed 500 words (including examples, excluding references). They should clearly present the research questions, the approach adopted, the methodology, the data, and the expected results. Abstracts should follow the template provided on the [LLcD2026 website](#).

References

- Bary, C., Berck, P. & Hendrickx, I. (2017). A memory-based lemmatizer for Ancient Greek. *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage*, 91–95.
- Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6: 213–234.
- Farina, A., Ciletti, M., McGillivray, B., & Ballatore, A. (2026). Sense-Based Annotation of Geographical Nouns in Ancient Greek and Latin: A Diachronic Study with LLMs. In *Proceedings of the 10th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature 2026*, 266–279, Rabat, Morocco.
- Fellbaum, C. (2006). WordNet and wordnets. In Goodman, K. S. & Goodman, Y. M. (eds.), *Encyclopedia of Language and Linguistics*. Elsevier, 2–665.
- Ghinassi, I., Tedeschi, S., Marongiu, P., Navigli, R. & McGillivray, B. (2024). Language pivoting from parallel corpora for word sense disambiguation of historical languages: A case study on Latin. In: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 10073–10084.
- Glynn, D. (2009). Polysemy, syntax, and variation: A usage-based method for cognitive semantics. In Evans, V. & Pourcel, S. (eds.), *New Directions in Cognitive Linguistics*. Amsterdam & Philadelphia: Benjamins, 77–106.
- Gulordava, K. & Baroni, M. (2011). A distributional similarity approach to the detection of semantic change in the Google Books Ngram corpus. *Proceedings of the GEMS 2011 Workshop on Geometrical Models of Natural Language Semantics*, 67–71. Liétard, B., Keller, M. & Denis, P. (2024). Towards an onomasiological study of lexical semantic change through the induction of concepts. *Proceedings of the 5th Workshop on Computational Approaches to Historical Language Change*, 158–167.
- Nwagwu, W. (2022). The rise and rise of natural language processing research, 1958–2021. In Review. <https://doi.org/10.21203/rs.3.rs-2265814/v1> (last access: 10 April 2026).
- Periti, F. & Tahmasebi, N. (2024). A systematic comparison of contextualized word embeddings for lexical semantic change. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2024)*, 4262–4282.
- Piotrowski, M. (2012). *Natural Language Processing for Historical Texts*. Springer International Publishing.
- Vatri, A. & McGillivray, B. (2020). Lemmatization for Ancient Greek: An experimental assessment of the state of the art. <https://doi.org/10.17863/CAM.55817> (last access: 10 April 2026).

Contact information

Andrea Farina

King's College, London (UK)
andrea.farina@kcl.ac.uk

Zinaida Geylikman

Independent Researcher (France)
zinaida.geylikman@gmail.com

Bastien Liétard

King's College, London (UK)
bastien.lietard@kcl.ac.uk

Barbara McGillivray

King's College, London (UK)
barbara.mcgillivray@kcl.ac.uk