NLP tools and methods for diachronic lexical semantics

Andrea Farina: andrea.farina@kcl.ac.uk, King's College London (United Kingdom)

Zinaida Geylikman: zinaida.geylikman@gmail.com, Independent Researcher (France)

Bastien Liétard: bastien.lietard@outlook.fr, King's College London (United Kingdom)

Barbara McGillivray: barbara.mcgillivray@kcl.ac.uk, King's College London (United Kingdom)

Subject

The workshop aims to gather contributions on NLP tools and methods for the diachronic study of lexical semantics in historical languages. By bringing together recent advances – from traditional approaches to BERT-based models and large language models – we seek to assess the current state of the art in the field. The focus is on diachrony across different languages and on lexical semantics broadly conceived, addressing two main questions: (1) does the performance of NLP tools vary depending on the historical period and language considered? (2) in which areas do NLP methods still require refinement, and where are manual or semi-automatic annotations more reliable than fully automatic approaches?

State of the art

Over the past decades, Natural Language Processing (NLP) has profoundly transformed linguistic research. From its origins in the 1950s, the field moved toward statistical methods in the 2000s and, more recently, toward deep learning and transformer-based architectures (Nwagwu, 2022; Lo et al., 2023). Tasks such as lemmatization, part-of-speech tagging, and syntactic parsing can now be performed automatically and at scale with high accuracy. This development also extends to historical and modern languages, for which increasingly reliable tools now support lemma identification and syntactic annotation within standardized frameworks such as Universal Dependencies (UD).

In lexical semantics, corpus-based approaches have long been integrated into multiple theoretical frameworks. Distributional semantics models meaning through contextual similarity (Boleda, 2020), while lexical-semantic networks such as WordNet (Fellbaum, 2006) provide structured representations of semantic relations. These methods have proven effective in identifying synonymy (Mikolov et al., 2013) and detecting diachronic semantic shifts through changes in distributional patterns (Gulordava & Baroni, 2011; Law, 2023). In parallel, Linked Open Data initiatives have enabled the construction of interoperable ontologies connecting lexical resources across datasets and languages. Together, these approaches allow large-scale modeling of semantic relations and change, though interpretative validation by human experts often remains necessary.

Research on historical languages has particularly benefited from recent advances in deep learning and the growing availability of structured corpora. Foundational work has focused on token-level tasks such as lemmatization, morphological analysis, and syntactic parsing (e.g., Piotrowski, 2012; Bary et al., 2017; Celano et al., 2016; Vatri & McGillivray, 2020; Favaro et al., 2023; Kupari et al., 2024; Alzetta & Montemagni, 2025; Gordin et al., 2025), with comparable efforts across a range of historical languages. More recently, large language models (LLMs) have been applied to translation, authorship attribution, spatial relation recognition, and increasingly to semantic analysis itself (e.g., Ghinassi et al. 2024; Farina et al., 2023; Schmidt et al., 2024; McGillivray and Farina, forth.). From a diachronic perspective, NLP tools and generative AI have also been employed to investigate metaphor and lexical semantic change (e.g. Periti and Tahmasebi, 2024; Liétard et al., 2024; Leotta et al., 2026; Colella & Farina, 2026).

Despite these promising results, significant challenges remain. Performance may vary depending on the historical period, genre, and language under consideration, and tasks requiring fine-grained semantic interpretation – such as usage-feature annotation in Cognitive Semantics (Glynn, 2009; 2010) – still often benefit from manual or semi-automatic approaches. Given the proliferation of electronic corpora, interoperable databases, and increasingly sophisticated AI systems, there is an urgent need to systematically evaluate how lexical meaning, variation, and diachronic change can be modeled reliably across large datasets, and to determine where automated methods require refinement or closer integration with expert-driven annotation.

Research Questions

This workshop seeks to investigate whether current NLP models and tools can support the annotation of lexical semantic features (including word senses, metaphors, and distinction between concrete and abstract meanings). In particular, it aims to assess under which conditions high-quality annotation can be achieved and how this may vary depending on several factors: the theoretical framework adopted (e.g. cognitive, frame, distributional, or structural semantics); the research objective (theoretical-semantic analysis vs. lexicographical output); the language under investigation; the chronological dimension (ancient *vs*. modern languages); the nature of the dataset (large or small corpora, written or multimodal data); and the specific computational approach employed, including established automatic annotation tools and generative AI systems. A systematic examination of these variables will allow the workshop to clarify both the potential and the current limitations of NLP-based approaches to diachronic lexical semantics.

Objectives

The primary objective of this workshop is to bring together researchers from diverse theoretical backgrounds who work on different languages and chronological periods within a diachronic perspective, in order to assess whether the current state of computational research enables reliable and efficient automatic annotation of lexical meaning. We particularly welcome contributions with a comparative, evaluative, or methodologically reflective perspective that systematically compare human-assisted annotation with fully automatic approaches.

We propose a dedicated workshop rather than a general session because the topic is at once highly specialized and broadly interdisciplinary. Scholars working on computational approaches to lexical semantics are often situated in different research communities – ranging from theoretical linguistics and lexicography to NLP and AI – and may not otherwise have a

focused venue for methodological exchange. The creation of a space specifically devoted to diachronic lexical semantics and computational methods aims to foster dialogue across frameworks and languages, encourage the sharing of tools and evaluation practices, and lay the groundwork for a collaborative network of researchers. A central goal is to identify the current strengths and weaknesses of computational approaches to lexical meaning and to explore collectively how existing limitations can be addressed.

Relevant (sub-)discipline(s)

Lexical Semantics, Historical Linguistics, Lexicology, Lexicography, Corpus Linguistics, NLP.

References

Alzetta, C., & Montemagni, S. (2025). Low- vs High-level Lemmatization for Historical Languages. A Case study on Italian. *Proceedings of CLiC-It 2025: Eleventh Italian Conference on Computational Linguistics, September 24 — 26, 2025, Cagliari, Italy.* https://clic2025.unica.it/wp-content/uploads/2025/09/3_main_long.pdf

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. https://doi.org/10.1145/3078081.3078100

Boleda, Gemma. 2020. Distributional Semantics and Linguistic Theory. Annual Review of Linguistics 6.213–234. https://doi:10.1146/annurev-linguistics-011619-030303.

Celano, G. G. A., Crane, G., & Majidi, S. (2016). Part of Speech Tagging for Ancient Greek. *Open Linguistics*, 2(1). https://doi.org/10.1515/opli-2016-0020

Colella, F., & Farina, A. (2026). Wounds and Strikes: Tracking Semantic Change in Latin vulnus and plaga with LLMs. *Quantitative Diachronic Linguistics and Cultural Analytics: Data-Driven Insights into Language and Cultural Change*. King's College London, 15–16 January 2026. London (United Kingdom).

Farina, A., Ghinassi, I., & McGillivray, B. (2023). Preliminary experiments on using Generative Large Language Models for Latin Word Sense Disambiguation. *Zenodo*. https://doi.org/10.5281/ZENODO.10298273

Favaro, M., Biffi, M., & Montemagni, S. (2023). POS Tagging and Lemmatization of Historical Varieties of Languages. The Challenge of Old Italian. *Italian Journal of Computational Linguistics*, 9(2). https://doi.org/10.4000/ijcol.1325

Fellbaum, C. 2006. WordNet and wordnets. In K. S. Goodman & Y. M. Goodman, *Encyclopedia of Language and Linguistics, 2–665*. Elsevier.

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),* pages 10073–10084, Torino, Italia. ELRA and ICCL.

Glynn, Dylan. (2009). Polysemy, syntax, and variation: A usage-based method for Cognitive Semantics. In V. Evans, & S. Pourcel (Eds.), *New directions in Cognitive Linguistics* (pp. 77–106). Amsterdam & Philadelphia: John Benjamins.

Glynn, Dylan. (2010). Testing the hypothesis: Objectivity and verification in usage-based Cognitive Semantics. In D. Glynn, & K. Fischer (Eds.), *Quantitative Cognitive Semantics:*

Corpus-driven approaches (pp. 239–270). Berlin & New York: Mouton de Gruyter. DOI: 10.1515/9783110226423

Gordin, S., Sahala, A., Spencer, S., & Klein, S. (2025). EvaCun 2025 Shared Task: Lemmatization and Token Prediction in Akkadian and Sumerian using LLMs. *Proceedings of the Second Workshop on Ancient Language Processing*, 242–250. https://doi.org/10.18653/v1/2025.alp-1.33

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 Semantics67–71 Stroudsburg*, PA: Assoc. Comput. Linguist

Kupari, H.-M. K., Henriksson, E., Laippala, V., & Kanerva, J. (2024). Improving Latin Dependency Parsing by Combining Treebanks and Predictions. *Proceedings of the 4th International Conference on Natural Language Processing for Digital Humanities*, 216–228. https://doi.org/10.18653/v1/2024.nlp4dh-1.21

Law, James. 2023. Regularity of Semantic Change in Romance Anatomical Terms. *Journal of Historical Linguistics* 13.295–325. doi:10.1075/jhl.21046.law.

Leotta, R., Lunardi, V., & Farina, A. (2026). Synsets, Emotions, and Cultural Shifts: A Diachronic Study of Conceptual Metaphors in Latin. *Quantitative Diachronic Linguistics and Cultural Analytics: Data-Driven Insights into Language and Cultural Change.* King's College London, 15–16 January 2026. London (United Kingdom).

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. https://doi.org/10.18653/v1/2024.lchange-1.15

Lo, A. W., Singh, M., & ChatGPT. (2023). From ELIZA to ChatGPT: The Evolution of Natural Language Processing and Financial Applications. *The Journal of Portfolio Management*, 49(7), 201–235. https://doi.org/10.3905/jpm.2023.1.512

McGillivray, B., & Farina, A. (Forth). Are Large Language Models able to grasp Latin semantics? A study on motion verbs. *Proceedings of the 23rd International Colloquium on Latin Linguistics. June 9-13, 2025, Udine (Italy).*

Mikolov T, Yih W, Zweig G 2013b. Linguistic regularities in continuous space word representations. *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2013) 746–51 Stroudsburg, PA: Assoc. Comput. Linguist.*

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

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)*, Mexico City, Mexico. Assoc. Comput. Linguist., 4262–4282. https://aclanthology.org/2024.naacl-long.240/

Piotrowski, M. (2012). *Natural Language Processing for Historical Texts*. Springer International Publishing. https://doi.org/10.1007/978-3-031-02146-6

Schmidt, G., Gorovaia, S., & Yamshchikov, I. P. (2024). Sui Generis: Large Language Models for Authorship Attribution and Verification in Latin (Version 1). *arXiv*. https://doi.org/10.48550/ARXIV.2410.09245

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

Short abstracts

Using computational semantic methods to explore onomasiological differences in the use of ancient scientific terminology

Alek Keersmaekers, Martijn Windelinckx, Toon Van Hal (KU Leuven)

Unlike modern scientific terminology, which is typically based on Greek or Latin roots, ancient Greeks had to rely on their native language to express scientific concepts. In the scholarship on Ancient Greek scientific language, the distinction between 'scientific' and 'everyday' language is often contested (Fögen 2015), but typically no quantitative evidence is provided. In this paper, we will make use of a large, automatically semantically annotated corpus (GLAUx, Keersmaekers 2021) to investigate the extent to which Ancient Greek had specialized technical vocabularies. Through the use of word sense disambiguation and distributional semantic modeling, we will explore in which situations more 'technical' and more 'everyday' terms exist to express similar semantic concepts, and how much variation exists within the realm of 'technical' vocabulary itself.

References

Fögen, Thorsten. 2015. Technical Literature. In Martin Hose & David Schenker (eds.), A Companion to Greek Literature, 266–279. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118886946.ch17.

Keersmaekers, Alek. 2021. The GLAUx corpus: methodological issues in designing a long-term, diverse, multi-layered corpus of Ancient Greek. In Nina Tahmasebi, Adam Jatowt, Yang Xu, Simon Hengchen, Syrielle Montariol & Haim Dubossarsky (eds.), Proceedings of the 2nd International Workshop on Computational Approaches to Historical Language Change 2021, 39–50. Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.lchange-1.6.

Assessing Semantic Disambiguation in Medieval French with HeidelBERT

Ragini Menon, Sabine Tittel (Heidelberg Academy of Sciences and Humanities)

Medieval French underwent profound linguistic change over the course of centuries, presenting significant challenges for computational modelling—challenges compounded by its status as an under-resourced historical language. Within the framework of the ALMA project ('Knowledge Networks in Medieval Romance Speaking Europe') and to investigate the semantic evolution specifically of medical and anatomical terminology in Old- and Middle French texts produced in France and England between the 12th and 15th centuries, we create a large, domain-specific textual corpus. Its texts undergo extensive annotation: part of speech-tagging, lemmatisation, and semantic disambiguation, the latter being by far the most challenging task. To accelerate the annotation, we develop HeidelBERT, a Large Language Model (LLM) for Old and Middle French. The model is trained on ALMeD, a fine-grained, genre-specific gold-standard dataset of semantically annotated medical literature with

approximately 36,000 tokens, covering three texts between the second quarter of the 13th c. and the second third of the 15th c. including the Anglo-Norman and Lorraine scriptae. HeidelBERT addresses key challenges in historical NLP, including data scarcity and semantic ambiguity, through the evaluation and adaptation of current state-of-the-art methodologies. To evaluate its performance in semantic disambiguation, we construct a small, dedicated test set of manually annotated ambiguous lexical items, and compare HeidelBERT's predictions to the reference annotations using quantitative metrics such as accuracy and F1-score. This will be complemented by a qualitative error analysis that examines systematic misclassifications, performance on low-frequency senses, and cases involving diachronic semantic shift – which will provide an assessment of both, overall accuracy and the model's sensitivity to historically complex variation.

Leveraging Large Language Models for Lexical Disambiguation of diachronic Italian within the LiITA (Linking Italian) Framework

Eleonora Litta, Giovanni Moretti, Marco Passarotti (Università Cattolica del Sacro Cuore)

The recent expansion of the Linked Open Data (LOD) paradigm has increasingly demonstrated how fundamental it is to link resources together to unlock the full potential of digital scholarship and data interoperability. Within this framework, linking historical language varieties to structured knowledge bases remains a significant challenge in computational linguistics due to linguistic drift and the scarcity of annotated diachronic data. This research presents a novel experimental pipeline designed to link MIDIA (Morfologia dell'Italiano in DIAcronia, http://www.corpusmidia.unito.it), a corpus of diachronic Italian material, fully annotated with lemma and part-of-speech information, to the LiITA (Linking Italian, http://www.liita.it) Knowledge Base. Specifically, we address the task of disambiguating 1:N lemma-POS matches, where a single token in the corpus maps to multiple potential candidates in the LiLa Lemma Bank, a LOD curated collection of citation forms (lemmas) that serves as a foundational "pivot" for Italian linguistic resources, given a set of dictionary definitions linked to each lemma. We propose a bespoke disambiguation pipeline that leverages the contextual reasoning capabilities of Large Language Models (LLMs) to interpret the semantic and syntactic nuances of historical Italian, effectively selecting the correct LiITA identifier from a set of candidates. To assess the robustness and generalisability of the model, we evaluate the pipeline across a representative sample of four out of seven distinct genres contained in the MIDIA corpus: literary, scientific, legal, and epistolary. This cross-genre evaluation allows us to measure how domain-specific terminology and stylistic variations impact the LLM's performance in the linking process, by providing an automated, scalable method for enriching diachronic corpora with linked open data, facilitating deeper longitudinal studies of the Italian language.

Capturing Historical Food Practices with LLMs: A Cross-Linguistic Study

Teresa Paccosi, Gauri Bhagwat, Marieke van Erp (DHLab, KNAW Humanities Cluster, Amsterdam)

This study presents an annotation scheme for English and Dutch aimed at identifying food and its historically relevant uses beyond mere consumption, including medicinal applications and preservation techniques. Based on guidelines inspired by Frame Semantics, we have created a high-quality reference corpus to support the future training of a classifier capable of automatically extracting different food identities in context, enabling large-scale quantitative analyses. We evaluate the potential for automating this annotation process by comparing human annotations with outputs generated by several open source LLMs. The models were prompted using a strategy combining Few-Shot learning with Chain-of-Thought reasoning to produce annotations, directly comparable with human annotations. Evaluation was conducted on 200 sentences for English and 200 for Dutch from household and recipe books, and medical-related texts. We also tested the impact of input length by comparing annotations generated from three-sentence segments versus the entire text. Results show good inter-annotator agreement and encouraging LLM performance, particularly for food entity recognition, while preservation-related frames remained more challenging. Shorter input segments result in improved model performance, highlighting the potential of this setting for semi-automatic annotation pipelines for this task.

Polysemy Across Time: LLM-Based Classification of Latin Modal Markers

Paola Marongiu (Institute for Computational Linguistics 'A. Zampolli', National Research Council, Pisa, Italy)

Modal markers are well known to undergo diachronic semantic shifts, often moving between different types of possibility and necessity (dynamic, deontic and epistemic; see Nuyts 2016; Dell'Oro 2023), and non-modal uses. Such pathways have been widely documented in cross-linguistic research. A major reference is the cross-linguistic modal map proposed by van der Auwera and Plungian (1998), which models relations between premodal, modal and postmodal meanings in synchrony and diachrony. Building on this framework, Magni (2005, 2010) developed a comprehensive map specific to Latin, tracing the modal pathways of markers such as debeo 'I have to', licet 'it is possible/allowed', necesse (est), oportet 'it is necessary' and possum 'I am able/I can'. For Latin, recently the WoPoss project has released a diachronic corpus annotated with modality (Dell'Oro 2022), and a dataset of diachronic modal maps based on lexicographic sources (Marongiu & Dell'Oro 2021). However, the extent to which LLMs can detect such meaning shifts remains unexplored.

This paper investigates whether contemporary LLMs can identify and classify semantic shifts in Latin modal markers, focusing on the distinction between modal and non-modal readings and on shifts across modal subtypes. We select a sub-set of a high-frequency markers (debeo, licet, possum) and compare several state-of-the-art LLMs in two settings: (i) zero-shot prompting and (ii) few-shot prompting enriched with structured information derived from the diachronic modal maps. Tasks include distinguishing modal from non-modal uses, assigning broad modal types (dynamic, deontic, epistemic), and testing sensitivity to the chronological emergence of meanings.

Based on recent work on Latin (e.g. McGillivray & Farina 2026; Ghizzota et al. 2026), we expect LLMs to capture coarse distinctions (modal/non-modal, possibility/necessity) but to struggle with fine-grained contextual readings; incorporating structured diachronic information in few-shot settings may improve alignment with gold annotations.

References

van der Auwera, Johan, and Vladimir A. Plungian. 1998. “Modality’s semantic map”. Linguistic Typology 2 (1): 79–124.

Dell’Oro, Francesca (dir.) 2022. WoPoss. A corpus to analyse the evolution of modality in the diachrony of the Latin language. Swiss National Science Foundation. https://woposs.unine.ch/.

Dell’Oro, Francesca. 2023. WoPoss guidelines for annotation. Revised version (Version v4). University of Neuchâtel – Swiss National Science Foundation. doi:10.5281/zenodo.10427053.

Ghizzota, E., Marongiu, P., Basile, P., Ferilli, S., McGillivray, B. 2026 (accepted/in press). Linguistic Knowledge Graphs for Sense Prediction: a case-study on Latin. LREC-COLING, Palma de Mallorca, Spain.

Magni, Elisabetta. 2005. “Modality’s semantic maps. An investigation of some Latin modal forms”. Journal of Latin Linguistics 9 (1): 325–336.

— . 2010. “Mood and modality”. In New perspectives on historical Latin syntax, ed. by Philip Baldi and Pierluigi Cuzzolin, 2:193–275. Berlin: Mouton de Gruyter.

Marongiu, Paola, and Francesca Dell’Oro. 2022. “From static to interactive maps: Drawing diachronic maps of (Latin) modality with Pygmalion”. Journal of Open Humanities Data 8

(2): 1–14. doi:10.5334/johd.58.

McGillivray, Barbara & Farina, Andrea. 2026. (Accepted/In press) Proceedings of the 23rd International Colloquium on Latin Linguistics.

Nuyts, Jan. 2016. “Analyses of the modal meanings”. In The Oxford handbook of modality and mood, ed. by Jan Nuyts and Johan van der Auwera, 31–49. Oxford, UK: Oxford University Press.