



LREC 2026

NEOLOGY AND LARGE LANGUAGE MODELS

Workshop Proceedings

Editors

Valunaite Oleskeviciene Giedre and Giouli Voula

16 MAY 2026

©ELRA Language Resources Association (ELRA), 2026

These proceedings are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-70-8

EAN 9782493814708

Preface

Neology, the process of creating and disseminating new words, has a long tradition in lexicography, lexicology, corpus and computational linguistics, and sociolinguistics. Recently, with the advent and rapid diffusion of Large Language Models (LLMs), the analysis of lexical innovation has entered a new era. LLMs process linguistic data at an unprecedented scale, modeling language in ways that mirror, extend, and sometimes distort human lexical creativity (Qin et al., 2025; Hosseini-Kivanani, 2025). Undoubtedly, LLMs are powerful tools for detecting new words (Tomaszewska et al., 2025) and tracking their life cycle (Cartier, 2017) across domains, genres, and communities of use (Ichien et al., 2024), thus complementing and extending corpus-based approaches to neology. Moreover, LLMs not only absorb neologisms from their training data but also have the capacity to generate novel lexical items, metaphors, or hybrid forms in response to prompts (Iwamoto and Kanayama, 2024). This raises important questions about authority, legitimacy, and the mechanisms of linguistic innovation. From a sociolinguistic perspective, the interaction between LLMs and neology highlights both opportunities and risks. On one hand, LLMs could assist communities in documenting and systematizing emergent lexical forms, thus supporting language revitalization. On the other hand, their biases and uneven coverage may privilege neologisms in high-resource languages while under-representing or distorting innovation in smaller linguistic communities. Ultimately, this dynamic positions LLMs not just as tools for text processing but as active participants in shaping the future trajectories of language use and innovation. These developments open new perspectives for tracking linguistic creativity, modeling semantic shifts, and supporting lexicographic and terminological research.

The NeoLLM2026 workshop explores the intersection of LLMs and neology, focusing on—but not limited to—how cutting-edge computational methodologies leveraging LLMs can identify, generate, analyze, and evaluate new words and semantic shifts across languages and language varieties.

The invited speaker, Maciej Ogrodniczuk, from the Institute of Computer Science of the Polish Academy of Sciences, in his talk **From "Neologism of the Week" to "Neologism of the Year"**, addresses how Large Language Models (LLMs) can facilitate the study of lexical innovation by shifting the focus of neology towards continuous, data-driven observation and how LLMs can influence the detection, interpretation, and legitimisation of new words in contemporary discourse. At the empirical core of the study lies NeoN: an LLM-enhanced system that automatically detects, monitors, and analyzes neologisms in Polish, with a particular focus on its 'neologism of the week' module. Combining frequency dynamics, contextual evidence, and automated definition and categorization, the system documents which new words appear in discourse and how they gain or lose relevance over time. Situating NeoN within a wider cross-disciplinary framework, the talk reflects on how LLM-based approaches can complement traditional lexicography and support linguistic documentation, offering new methodological pathways for studying linguistic change. The talk will also explore whether the 'neologism of the year' is merely a linguistic artifact or a socio-technical construct, exposing the evolving interplay between human creativity, media dynamics and artificial intelligence.

Organizing Committee

Florentina Armaselu (University of Luxembourg)
Voula Giouli (Aristotle University of Thessaloniki)
Barbara Lewandowska-Tomaszczyk (University of Applied Sciences in Konin)
Chaya Liebeskind (Jerusalem College of Technology)
Barbara McGillivray (King's College London)
Giedre Valunaite Oleskeviciene (Mykolas Romeris University)

Scientific Committee

Florentina Armaselu (University of Luxembourg, Luxembourg)
Giorgio Maria Di Nunzio (University of Padua, Italy)
Radovan Garabík (Ludovit Stur Institute of Linguistics, Slovak Republic)
Voula Giouli (Aristotle University of Thessaloniki)
Anas Fahad Khan (CNR-Istituto di Linguistica Computazionale "Antonio Zampolli," Italy)
Barbara Lewandowska-Tomaszczyk (University of Applied Sciences in Konin)
Chaya Liebeskind (Jerusalem College of Technology, Israel)
Elpida Loupaki (Aristotle University of Thessaloniki, Greece)
Barbara McGillivray (King's College London)
Liudmila Mockiene (Mykolas Romeris University, Lithuania)
Maciej Ogrodniczuk (Institute of Computer Science, Polish Academy of Sciences, Poland)
Atul Kr. Ojha (University of Galway, Ireland)
Ana Ostroški Anić (Institute of Croatian Language and Linguistics, Croatia)
Marko Robnik-Šikonja (University of Ljubljana, Slovenia)
Purificação Silvano (University of Porto, Portugal)
Enriketa Sogutlu (Beder University, Albania)
Ranka Stankovic (University of Belgrade, Serbia)
Giovanni Luca Tallarico (University of Verona, Italy)
Giedre Valunaite Oleskeviciene (Mykolas Romeris University, Lithuania)
Andrius Utkā (Vytautas Magnus University)
Federica Vezzani (University of Padua, Italy)

Table of Contents

<i>From 124 Million Tokens to 1,021 Neologisms: A Large-Scale Pipeline for Automatic Neologism Detection</i>	
Diego Rossini and Lonneke van der Plas	1
<i>High Resource Bias in AI-Driven Neology: Structural Inequality in Lexical Innovation</i>	
Wajdi Zaghouani	16
<i>Do LLMs Know What Luxembourgish Borrows? Probing Lexical Neology in Low-Resource Multilingual Models</i>	
Nina Hosseini-Kivanani	27
<i>Lexical Innovation in Business Colour Idioms: Evidence from Large Language Models in Five Languages</i>	
Giedre Valunaite Oleskeviciene, Ágnes Abuczki, Ganit Richter, Berat Ujkani, Vera Moitinho de Almeida and Pedro Madeira	39
<i>Where in Semantic Space Do Spanish Neologisms Emerge?</i>	
Bianca Delgado and Shira Wein	47
<i>Assessing the Pragmatic Competence of LLMs Regarding Novel Discourse Markers in Digital Communication</i>	
Ágnes Abuczki and Giedre Valunaite Oleskeviciene	53
<i>A Comparative Evaluation of Semantic Ambiguity Detection in Two LLMs</i>	
Lili Tamas	60
<i>LLM-Based Frame and Stance Annotation for 19th-Century Rumour Discourse in US and UK Newspapers</i>	
Wanshu Zhang	66

Workshop Program

Saturday, 5/16/2026

14:00–18:00	Session Workshop: Neology and Large Language Models Room: Room 9 Chairs: Giedre Valunaite Oleskeviciene, Voula Giouli
14:40–15:00	<i>From 124 Million Tokens to 1,021 Neologisms: A Large-Scale Pipeline for Automatic Neologism Detection</i> Diego Rossini and Lonneke van der Plas
15:00–15:20	<i>High Resource Bias in AI-Driven Neology: Structural Inequality in Lexical Innovation</i> Wajdi Zaghoulani
15:20–15:40	<i>Do LLMs Know What Luxembourgish Borrows? Probing Lexical Neology in Low-Resource Multilingual Models</i> Nina Hosseini-Kivanani
15:40–16:00	<i>Lexical Innovation in Business Colour Idioms: Evidence from Large Language Models in Five Languages</i> Giedre Valunaite Oleskeviciene, Ágnes Abuczki, Ganit Richter, Berat Ujkani, Vera Moitinho de Almeida and Pedro Madeira
16:30–16:50	<i>Where in Semantic Space Do Spanish Neologisms Emerge?</i> Bianca Delgado and Shira Wein
16:50–17:10	<i>Assessing the Pragmatic Competence of LLMs Regarding Novel Discourse Markers in Digital Communication</i> Ágnes Abuczki and Giedre Valunaite Oleskeviciene
17:10–17:30	<i>A Comparative Evaluation of Semantic Ambiguity Detection in Two LLMs</i> Lili Tamas
17:30–17:50	<i>LLM-Based Frame and Stance Annotation for 19th-Century Rumour Discourse in US and UK Newspapers</i> Wanshu Zhang

