



LREC 2026

**Natural Scientific Language Processing (NSLP 2026) @
LREC 2026**

Workshop Proceedings

Editors

**Georg Rehm, Stefan Dietze, Danilo Dessí,
Diana Maynard, Sonja Schimmmler**

12 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 978-2-493814-61-6

Preface

Scientific research is almost exclusively published in unstructured text formats, which are not readily machine-readable. While technological approaches can help to get this flood of scientific information and new knowledge under control, the development of such technologies is very complex in practice and hinders the creation of infrastructures and systems to track research and assist the scientific community with applications such as dedicated scientific search engines and recommender systems. The 3rd International Workshop on Natural Scientific Language Processing (NSLP 2026) aims to bring together researchers working on the processing, analysis, transformation and making-use-of scientific language including all relevant sub-topics. NSLP 2026 was a full-day workshop co-located with LREC 2026 held in Palma de Mallorca, Spain in May 2026.

In addition to the opportunity to submit papers covering original research that fits the workshop's topics of interest, the event also offered two shared tasks: Scientific Fact-Checking of Social Media Claims (ClimateCheck 2026) and Software Mention Disambiguation and Coreference Resolution (SOMD 2026). This proceedings volume contains two overview articles that describe these two shared tasks in more detail as well as several papers submitted to the two shared tasks.

With NSLP 2026 as the third edition of this workshop series, we were happy about a total of 38 submissions out of which 29 papers were accepted (76%) after a thorough, double-blind peer-review process with three reviews per submission. The NSLP 2026 workshop consisted of paper and poster presentations as well as two invited keynote presentations, given by Yufang Hou, ITU Austria, and Iryna Gurevych, TU Darmstadt, Germany. The abstracts of the two keynote presentations are included in this proceedings volume.

The workshop programme is structured into three sessions and a total of four tracks. Session 1 includes the two *Shared Task Overview Papers*. Session 2, *Scholarly Infrastructure and Data Foundations*, consists of two tracks. Track A, *Knowledge Graphs and Taxonomy*, builds what we can call the research or scientific knowledge graph, while Track B, *Domain-Specific Resources and Metadata*, populates the graph with data. Session 3, *Reliability, Reach and Scale*, concentrates on two tracks: Track C, *Factuality and Agentic Reasoning* asks if the data is true and Track D, *Multilingualism and Model Optimization*, asks how we scale that truth globally.

We would like to thank the LREC 2026 organisers and workshop chairs for accepting our workshop proposal. We would also like to extend our gratitude to our keynote speakers for their insightful talks. Thanks are also due to the members of the Programme Committee for reviewing the paper submissions under rather tight deadlines. Finally, we would like to thank Raia Abu Ahmad and Ekaterina Borisova for setting up and maintaining the workshop website.¹

This workshop was organised under the umbrella of the project NFDI for Data Science and Artificial Intelligence (NFDI4DS), which is part of the wider German NFDI initiative (National Research Data Infrastructure). Without the financial support of this project, the NSLP 2026 workshop would not have been possible.

Georg Rehm, Stefan Dietze, Danilo Dessí, Diana Maynard, Sonja Schimmler

¹<https://nfdi4ds.github.io/nslp2026/>

Organizing Committee

- Georg Rehm – Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI) & Humboldt-Universität zu Berlin, Germany
- Stefan Dietze – GESIS Leibniz Institut für Sozialwissenschaften, Cologne & Heinrich-Heine-Universität Düsseldorf, Germany
- Danilo Dessí – University of Sharjah, UAE
- Diana Maynard – University of Sheffield, UK
- Sonja Schimmmer – Technical University of Berlin & Fraunhofer FOKUS, Germany

Programme Committee

- Raia Abu Ahmad – Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany
- Marcel R. Ackermann – Schloss Dagstuhl, Leibniz Center for Informatics (LZI), DBLP, Germany
- Sören Auer – TIB Leibniz Information Centre for Science and Technology, Germany
- Fabio Barth – Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany
- Ekaterina Borisova – Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany
- Davide Buscaldi – LIPN, CNRS, University Paris 13, France
- Leyla Jael Castro – ZB MED Information Centre for Life Sciences, Germany
- Mathieu d'Aquin – Université de Lorraine, France
- Catherine Faron – Université Cote d'Azur, France
- Paul Groth – University of Amsterdam, The Netherlands
- Robert Jäschke – Humboldt-Universität zu Berlin, Germany
- Frank Krüger – Wismar University of Applied Sciences, Germany
- Julia Matela – Wismar University of Applied Sciences, Germany
- Philipp Mayr – GESIS Leibniz-Institute for the Social Sciences, Germany
- Julian Moreno-Schneider – Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Germany
- Pedro Ortiz Suarez – Common Crawl Foundation, USA
- Wolfgang Otto – GESIS Leibniz-Institute for the Social Sciences, Germany
- Haris Papageorgiou – R. C. Athena, Greece
- Harald Sack – FIZ Karlsruhe, Germany

- Angelo Salatino – The Open University, UK
- Philipp Schaer – TH Köln (University of Applied Sciences), Germany
- Vera Schmitt – Technische Universität Berlin, Germany
- Veronika Solopova – Technische Universität Berlin, Germany
- Sharmila Upadhyaya – GESIS Leibniz-Institute for the Social Sciences, Germany
- Max Upravitelev – Technische Universität Berlin, Germany
- Ricardo Usbeck – Leuphana University, Germany
- Aida Usmanova – Leuphana University, Germany
- Thanasis Vergoulis – R. C. Athena, Greece

Table of Contents

<i>AstroConcepts: A Large-Scale Multi-Label Classification Corpus for Astrophysics</i> Atilla Kaan Alkan, Felix Grezes, Sergi Blanco-Cuaresma, Jennifer Lynn Bartlett, Daniel Chivvis, Anna Kelbert, Kelly Lockhart and Alberto Accomazzi.....	1
<i>Benchmarking LLMs for ARR Area Assignment: Evidence and Implications for Assignment Strategies</i> Eileen Bingert, Diego Alves and Stefania Degaetano-Ortlieb.....	13
<i>Benchmarking Retrieval-Augmented Generation for Scientific Knowledge QA in European Portuguese</i> Jose Matos, Catarina Silva and Hugo Goncalo Oliveira.....	25
<i>Beyond Abstracts: A Biomedical MeSH Indexing Corpus Incorporating Summarized Methods Sections</i> Sujoy Datta, Robert E. Mercer and Xindi Wang.....	32
<i>Challenges and Opportunities for NSLP in Scientific Publishing – A Case Study</i> Thomas Kleinbauer, Michael Didas and Michael Wagner.....	44
<i>ClimateCheck 2026: Scientific Fact-Checking and Disinformation Narrative Classification of Climate-related Claims</i> Raia Abu Ahmad, Max Upravitelev, Aida Usmanova, Veronika Solopova and Georg Rehm.....	49
<i>ClimateSense at ClimateCheck 2026</i> Thibault Ehrhart, Gregoire Burel and Raphael Troncy.....	66
<i>Comparing LLM-Based Knowledge Graph Extraction Approaches on Literary Studies in Spanish: A Case Study on Orbis Tertius</i> Federico Cortes.....	78
<i>Demystifying Funding: Reconstructing a Unified Dataset of the UK Funding Lifecycle</i> William Thorne, Rupert Shepherd and Diana Maynard.....	88
<i>Do Lexical and Contextual Coreference Resolution Systems Degrade Differently under Mention Noise? An Empirical Study on Scientific Software Mentions</i> Atilla Kaan Alkan, Felix Grezes, Jennifer Lynn Bartlett, Anna Kelbert, Kelly Lockhart and Alberto Accomazzi.....	97
<i>Do We Need Bigger Models for Science? Task-Aware Retrieval with Small Language Models</i> Florian Kelber, Matthias Jobst, Yuni Susanti and Michael Färber.....	108
<i>EarlySciRev: A Dataset of Early-Stage Scientific Revisions Extracted from LaTeX Writing Traces</i> Léane Jourdan, Julien Aubert-Bédouchaud, Yannis Chupin, Marah Baccari and Florian Boudin.....	119
<i>Enhancing Factuality and Transparency in Generative Models for Biomedical Question Answering</i> Ankita Behura, Siting Liang and Daniel Sonntag.....	127

<i>Enhancing Scholarly Knowledge Graphs via Domain-Specific Entity Detection and Linking</i> Nicolau Duran-Silva, César A. Parra-Rojas, Pablo Accuosto, Julian Moreno-Schneider and Georg Rehm	140
<i>Evaluating Generative Large Language Models for Portuguese Scientific Information Extraction</i> Tomás Pinto, Catarina Silva and Hugo Goncalo Oliveira	155
<i>From Slides to Chatbots: Enhancing Large Language Models with University Course Materials</i> Tu Anh Dinh, Philipp Nicolas Schumacher and Jan Niehues	168
<i>Generating Research Data Metadata from Their Accompanying README Files</i> Kotaro Sekido, Yu Watanabe, Koichiro Ito and Shigeki Matsubara	180
<i>Identifying Implicit Research Data References in Paper Citations</i> Koshi Motegi, Koichiro Ito and Shigeki Matsubara	186
<i>Improving Completeness in Deep Research Agents through Targeted Enrichment</i> Jesse Wonnink, Jakub Zavrel and Paul Groth	193
<i>MioFFAn: An Annotation Software for Formula Formalization with LLM Automation Capabilities</i> Nicolas Sibuet Ruiz, Horacio Saggion and Riccardo Rossi	206
<i>Normalizing Section Names and Structure of Scientific Articles</i> Nicolau Duran-Silva, Julian Moreno-Schneider, César A. Parra-Rojas and Georg Rehm	218
<i>Retrieval-Augmented LLMs and Encoder Models for Multi-Label Climate Disinformation Narrative Classification</i> Neda Foroutan, Alexandra Tsiakalou and Vera Schmitt	225
<i>The Linguist's Lie Detector: Linguistic Knowledge in Large Language Models</i> Lucía Catalán Gris, Kim Gerdes and John S. Y. Lee	235
<i>The Software Mention Detection and Coreference Resolution Shared Task 2026</i> Sharmila Upadhyaya, Wolfgang Otto, Julia Matela, Frank Krüger and Stefan Dietze ..	247
<i>Towards Efficient Self-Explainable Climate-Related Claim Verification with Generative Models</i> Siting Liang, Omar Adjali and Daniel Sonntag	255
<i>Transferring Scientific English Pre-Trained Language Models to Multiple Languages Using Cross-Lingual Transfer</i> Nikolas Ching-Pu Rauscher, Fabio Barth and Georg Rehm	261
<i>Transformer Encoders with Heuristic-Guided Contrastive Learning for Software Coreference Resolution</i> Mahmoud Hassan and Dipendra Yadav	270
<i>UniCite: A Dataset and Unified Hierarchical Taxonomy for Multi-Dimensional Citation Analysis</i> Amina Mourky, Elena Leitner, Julian Moreno-Schneider, Raia Abu Ahmad, Ekaterina Borisova and Georg Rehm	277
<i>XplaiNLP @ ClimateCheck 2026 Task 2: Comparing Hierarchical Approaches for Fine-Grained Climate Disinformation Narrative Classification</i> Arthur Hilbert, Jing Yang and Vera Schmitt	289

Workshop Program

09:00-09:10	Welcome and Opening
09:10-09:50	Invited Keynote Speech – Yufang Hou
	Session 1: Shared Task Overview Papers
09:50-10:10	<i>ClimateCheck 2026: Scientific Fact-Checking and Disinformation Narrative Classification of Climate-related Claims</i> Raia Abu Ahmad, Max Upravitelev, Aida Usmanova, Veronika Solopova and Georg Rehm
10:10-10:30	<i>The Software Mention Detection and Coreference Resolution Shared Task 2026</i> Sharmila Upadhyaya, Wolfgang Otto, Julia Matela, Frank Krüger and Stefan Dietze
10:30-11:00	Coffee Break and Poster Session
	Session 2: Scholarly Infrastructure and Data Foundations
	Track A: Knowledge Graphs and Taxonomy
11:00-11:20	<i>Comparing LLM-Based Knowledge Graph Extraction Approaches on Literary Studies in Spanish: A Case Study on Orbis Tertius</i> Federico Cortes
11:20-11:40	<i>Enhancing Scholarly Knowledge Graphs via Domain-Specific Entity Detection and Linking</i> Nicolau Duran-Silva, César A. Parra-Rojas, Pablo Accuosto, Julian Moreno-Schneider and Georg Rehm
11:40-12:00	<i>UniCite: A Dataset and Unified Hierarchical Taxonomy for Multi-Dimensional Citation Analysis</i> Amina Mourky, Elena Leitner, Julian Moreno-Schneider, Raia Abu Ahmad, Ekaterina Borisova and Georg Rehm
	Track B: Domain-Specific Resources and Metadata
12:00-12:20	<i>Beyond Abstracts: A Biomedical MeSH Indexing Corpus Incorporating Summarized Methods Sections</i> Sujoy Datta, Robert E. Mercer and Xindi Wang
12:20-12:40	<i>AstroConcepts: A Large-Scale Multi-Label Classification Corpus for Astrophysics</i> Atilla Kaan Alkan, Felix Grezes, Sergi Blanco-Cuaresma, Jennifer Lynn Bartlett, Daniel Chivvis, Anna Kelbert, Kelly Lockhart and Alberto Accomazzi
12:40-12:50	<i>Normalizing Section Names and Structure of Scientific Articles</i> Nicolau Duran-Silva, Julian Moreno-Schneider, César A. Parra-Rojas and Georg Rehm
12:50-13:00	<i>Generating Research Data Metadata from Their Accompanying README Files</i> Kotaro Sekido, Yu Watanabe, Koichiro Ito and Shigeki Matsubara

13:00-14:00	Lunch Break
	Session 3: Reliability, Reach and Scale Track C: Factuality and Agentic Reasoning
14:00-14:20	<i>Enhancing Factuality and Transparency in Generative Models for Biomedical Question Answering</i> Ankita Behura, Siting Liang and Daniel Sonntag
14:20-14:40	<i>The Linguist's Lie Detector: Linguistic Knowledge in Large Language Models</i> Lucía Catalán Gris, Kim Gerdes and John S. Y. Lee
14:40-15:00	<i>Improving Completeness in Deep Research Agents through Targeted Enrichment</i> Jesse Wonnink, Jakub Zavrel and Paul Groth
	Track D: Multilingualism and Model Optimization
15:00-15:20	<i>Transferring Scientific English Pre-Trained Language Models to Multiple Languages Using Cross-Lingual Transfer</i> Nikolas Ching-Pu Rauscher, Fabio Barth and Georg Rehm
15:20-15:40	<i>Evaluating Generative Large Language Models for Portuguese Scientific Information Extraction</i> Tomás Pinto, Catarina Silva and Hugo Goncalo Oliveira
15:40-16:00	<i>Do We Need Bigger Models for Science? Task-Aware Retrieval with Small Language Models</i> Florian Kelber, Matthias Jobst, Yuni Susanti and Michael Färber
16:00-16:30	Coffee Break and Poster Session
16:30-17:15	Invited Keynote Speech – Iryna Gurevych
17:15-17:30	Closing and end of workshop

Poster Session

Benchmarking LLMs for ARR Area Assignment: Evidence and Implications for Assignment Strategies

Eileen Bingert, Diego Alves and Stefania Degaetano-Ortlieb

Benchmarking Retrieval-Augmented Generation for Scientific Knowledge QA in European Portuguese

Jose Matos, Catarina Silva and Hugo Goncalo Oliveira

Challenges and Opportunities for NSLP in Scientific Publishing – A Case Study

Thomas Kleinbauer, Michael Didas and Michael Wagner

ClimateSense at ClimateCheck 2026

Thibault Ehrhart, Gregoire Burel and Raphael Troncy

Demystifying Funding: Reconstructing a Unified Dataset of the UK Funding Lifecycle

William Thorne, Rupert Shepherd and Diana Maynard

Do Lexical and Contextual Coreference Resolution Systems Degrade Differently under Mention Noise? An Empirical Study on Scientific Software Mentions

Atilla Kaan Alkan, Felix Grezes, Jennifer Lynn Bartlett, Anna Kelbert, Kelly Lockhart and Alberto Accomazzi

EarlySciRev: A Dataset of Early-Stage Scientific Revisions Extracted from LaTeX Writing Traces

Léane Jourdan, Julien Aubert-Bédouchaud, Yannis Chupin, Marah Baccari and Florian Boudin

From Slides to Chatbots: Enhancing Large Language Models with University Course Materials

Tu Anh Dinh, Philipp Nicolas Schumacher and Jan Niehues

Identifying Implicit Research Data References in Paper Citations

Koshi Motegi, Koichiro Ito and Shigeki Matsubara

MioFFAn: An Annotation Software for Formula Formalization with LLM Automation Capabilities

Nicolas Sibuet Ruiz, Horacio Saggion and Riccardo Rossi

Retrieval-Augmented LLMs and Encoder Models for Multi-Label Climate Disinformation Narrative Classification

Neda Foroutan, Alexandra Tsiakalou and Vera Schmitt

Towards Efficient Self-Explainable Climate-Related Claim Verification with Generative Models

Siting Liang, Omar Adjali and Daniel Sonntag

Transformer Encoders with Heuristic-Guided Contrastive Learning for Software Coreference Resolution

Mahmoud Hassan and Dipendra Yadav

XplaiNLP @ ClimateCheck 2026 Task 2: Comparing Hierarchical Approaches for Fine-Grained Climate Disinformation Narrative Classification

Arthur Hilbert, Jing Yang and Vera Schmitt

Invited Keynote

Yufang Hou

ITU Austria, Austria

Synthesizing Scientific Knowledge: From Biomedical Evidence to NLP Claims

Abstract

The exponential growth of scholarly literature poses a substantial challenge for researchers seeking to stay current with the latest findings and synthesize knowledge effectively. In this talk, I will first present our recent studies on supporting domain experts in synthesizing biomedical research findings, guided by established evidence synthesis principles for integrating and interpreting potentially biased and conflicting scientific evidence. Next, I will introduce our work on content meta-analysis in NLP literature, including building leaderboards and tracking the evolution of scientific claims over time. Finally, I will discuss open research challenges in modeling and reasoning over scientific knowledge encoded in scholarly documents.

Bio

Yufang Hou is a professor at IT:U – Interdisciplinary Transformation University Austria. At IT:U, she leads the NLP group with a strong focus on large language model (LLM) governance, computational argumentation, fact-checking, knowledge representation and reasoning, and human-centered NLP applications in science, health, and education. Yufang has served as an organizing committee member and senior program committee member for various NLP conferences, and has co-organized several workshops on topics such as argument mining, efficient NLP, and AI for Science.

Invited Keynote

Iryna Gurevych

Technical University of Darmstadt, Germany

Welcoming AI as a New Colleague: How Should We Evaluate AI for Science?

Abstract

AI is reshaping the work of researchers in real time. With the emergence of tools such as AI Scientist in 2024, a new era of AI-driven scientific discovery has begun. At the same time, high-profile venues are beginning to experiment with AI-assisted peer review. Yet the core question remains unresolved: How should we meaningfully evaluate AI for scientific work? In this talk, I will highlight challenges encountered in several projects focused on AI-assisted scientific communication and discuss how these challenges inform evaluation practices. I will examine methods for assessing AI-generated related-work sections and explore how automated reviewers can be evaluated with respect to research-design reasoning and novelty assessment. Together, these examples illustrate both the promise and the complexity of evaluating AI as it increasingly acts as a collaborator in scientific communication.

Bio

Iryna Gurevych is Professor of Ubiquitous Knowledge Processing in the Department of Computer Science at the Technical University of Darmstadt in Germany. She also is an adjunct professor at MBZUAI in Abu-Dhabi, UAE, and an affiliated professor at INSAIT in Sofia, Bulgaria. She is widely known for fundamental contributions to natural language processing (NLP) and machine learning. Professor Gurevych is a past president of the Association for Computational Linguistics (ACL), the leading professional society in NLP. Her many accolades include being a Fellow of the ACL, an ELLIS Fellow, and the recipient of an ERC Advanced Grant. Most recently, she has received the 2025 Milner award of the British Royal Society for her major contributions to NLP and artificial intelligence that combine deep understanding of human language and cognitive faculty with the latest paradigms in machine learning.