LREC 2026

The Fourth Workshop on the Role of Resources in the Age of Large Language Models (RESOURCEFUL 2026) @ LREC 2026

Workshop Proceedings

Editors
Felix Morger, Nikolai Ilinykh, Barbara Scalvini, Simon Dobnik, Dana Dannélls

May 11, 2026

The Role of Resources in the Age of Large Language Models (RESOURCEFUL 2026)

The RESOURCEFUL organizers also gratefully acknowledge the support of Språkbanken Text, University of Gothenburg, EUTOPIA and CLARIN for supporting the workshop.

ISBN 978-2-493814-94-4

Message from the organisers

The fourth workshop on the role of resources in the age of large language models was held in Palma, Mallorca (Spain), on May 11, 2026. The workshop was conducted in person, while also providing an option for online participation.

This was the fourth edition of the workshop and, in line with the goals of the previous three workshops and current trends in linguistics, computational linguistics, and natural language processing, RESOURCEFUL-2026 focused on the role of resources in the age of large language models (LLMs). The workshop continued the RESOURCEFUL series by addressing how the language resources community can respond to the methodological, ethical, and practical challenges introduced by LLMs.

The language resources community has long provided the empirical foundation for language technology by building datasets that have been crucial for the development of NLP models. However, the introduction of LLMs, trained on vast and often undisclosed text collections, has disrupted this ecosystem. Traditional notions and methods of resource building are evolving: the boundaries between training and evaluation data are increasingly blurred, the very idea of truly "unseen" data is becoming harder to maintain, and synthetic linguistic data can now be generated at scale to support the creation of new resources. These developments raise fundamental questions about how we evaluate models, ensure data transparency, and preserve the integrity of linguistic resources. RESOURCEFUL-2026 therefore aimed to stimulate critical dialogue on data creation, authenticity, and representation in the age of LLMs.

The workshop aimed to bring together researchers involved in the creation, validation, and evaluation of next-generation language resources. In particular, it sought to promote discussion among traditional resource builders, evaluation specialists, linguists, field researchers, and LLM researchers, creating a shared forum to redefine the role of resources in NLP.

The call for papers for RESOURCEFUL-2026 invited contributions on the following topics:

- Novel approaches beyond static datasets; resources as processes; reusable, dynamic, and interactive resources
- Documentation, reproducibility, and transparency in procedurally generated or evolving resources
- Limitations and opportunities in using LLMs as "judges" or co-annotators to support expert-based linguistic annotation
- Quantifying linguistic, pragmatic, and cultural dimensions, and related biases, for resource creation including LLM-generated data
- Semi-automatic and human-in-the-loop methods for benchmark creation and model evaluation
- Synthetic and transfer-based methods for low-resource and domain-specific languages
- Evaluation under data scarcity, domain shift, or limited access to real data or annotators
- Maintaining and updating benchmarks in the LLM era
- Methods for generating and benchmarking synthetic linguistic data, and incorporating such data in model training and evaluation

- Purpose-based, Turing-test-inspired, or interaction-based evaluation of NLP systems
- Data ownership, governance, consent, and community-centered perspectives in data creation for under-represented languages
- Ethical and legal implications of automatically generated data
- Metadata and documentation practices for evolving and synthetic resources
- Long-term sustainability and openness of linguistic resources

We invited both archival and non-archival submissions. In total, **29** submissions were received, of which **18** were archival. The program committee (PC) consisted of **11** Programme Chairs and **37** reviewers. Based on the PC assessments regarding the content and quality of the submissions, the program chairs decided to accept **19** submissions for presentation. Together with the non-archival submissions, this resulted in a program consisting of **8** talks and **11** posters.

The accepted submissions addressed a broad range of topics related to the creation and evaluation of language resources in the age of LLMs. A strong focus of the programme was on benchmarking and evaluation, including work on machine translation, question answering, reasoning, hallucination, natural language generation, cultural alignment, pragmatics, and safety. Another prominent theme concerned resource creation and annotation methodologies, including LLM-assisted annotation, semi-automatic benchmark construction, and model-supported dataset development. The programme also highlighted multilinguality and linguistic diversity through work on low-resource, under-resourced, and historically grounded settings. Collectively, the contributions highlighted both the sustained importance of carefully curated linguistic resources and the need to rethink how such resources are created, validated, and maintained in light of LLM-based methods.

The themes emerging from both the workshop topic and the accepted submissions also motivated a panel discussion on the future of language resources in the age of LLMs. Among the issues expected to shape the discussion are the design and maintenance of benchmarks in the LLM era, the role of LLMs in annotation and resource creation, and the challenges of multilinguality, cultural grounding, and linguistic diversity.

The workshop had three keynote speakers: Mark Fišel (University of Tartu, Estonia), Maria Gavriilidou (Institute for Language and Speech Processing / RC Athena, Greece), and Tiago Torrent (Federal University of Juiz de Fora, Brazil). The workshop also had a panel discussion on the topics described above.

Words of appreciation and acknowledgment are due to the program committee, our sponsors, the local LREC 2026 organisers, and all authors and participants who contributed to making RESOURCEFUL 2026 a stimulating and successful event.

The RESOURCEFUL 2026 Program Chairs

Organizing Committee

Felix Morger, University of Gothenburg
Nikolai Ilinykh, University of Gothenburg
Barbara Scalvini, University of the Faroe Islands
Simon Dobnik, University of Gothenburg
Dana Dannélls, Språkbanken Text, University of Gothenburg
Beáta Megyesi, Dept. of Linguistics, Stockholm University
Bolette Sandford Pedersen, University of Copenhagen
Micaella Bruton, Dept. of Linguistics, Stockholm University
Dávid í Lág, University of the Faroe Islands, Faroe Islands
Alina Karakanta, Leiden University Centre for Linguistics
Joakim Nivre, Dept. of Linguistics and Philology, Uppsala University
Iben Nyholm Debess University of the Faroe Islands
Lilja Øvrelid Professor, Dept. of Informatics, University of Oslo
Sara Stymne, Dept. of Linguistics and Philology, Uppsala University
Jörg Tiedemann, Dept. of Digital Humanities, University of Helsinki, Finland
Crina Tudor, Dept. of Linguistics, Stockholm University

Programme Committee

Spela Arhar Holdt
Meriem Beloucif, Micaella Bruton, Lars Bungum
Pierluigi Cassotti
Amandine Decker
Hafsteinn Einarsson
Mariia Fedorova, Emilie Francis
Sardana Ivanova
Simeon Junker
Alina Karakanta
Erik Lagerstedt, Herb Lange
Celine Leuzinger, Anna Lindahl
Arianna Masciolini, John P. McCrae, Ricardo Muñoz Sánchez, Petter Mæhlum
Sanni Nimb, Joakim Nivre, Bill Noble, Nathalie Norman
Sussi Olsen
Danila Petrelli, Eva Pettersson
Michael Rießler
Annika Simonsen, Jákup Svøðstein, Maria Irena Szawerna, Tom Södahl Bladsjö
Joerg Tiedemann, Tiago Timponi Torrent, Crina Tudor
Thomas Vakili

Table of Contents

Workshop Program

09:00–09:05 **Welcome**

09:05–09:45 **Keynote I: Mark Fišel**

09:45–10:25 **Oral talks I**

09:45–10:05 *Lost in Translation: Repurposing semantic similarity benchmarks for evaluating lexical-semantic consistency in LLM-based machine translation*
Quin Ye and Jelke Bloem

10:05–10:25 *Bridging the Low Resource Gap in Historical Cryptology: A Multilingual Diachronic Synthetic Dataset for Reproducible Cryptanalysis*
Micaella Bruton, Meriem Beloucif and Beáta Megyesi

10:25–11:00 **Coffee break**

11:00–11:40 **Keynote II: Tiago Torrent**

11:40–12:40 **Oral talks II**

11:40–12:00 *Cultural Grounding in Swedish: Extending an Everyday Knowledge Benchmark for LLMs*
Meriem Beloucif and Johan Sjons

12:00–12:20 *Entity Linking for Faroese Using Large Language Models with Web Search*
Annika Simonsen, Iben Nyholm Debess and Hafsteinn Einarsson

12:20–12:40 *From Polyester Girlfriends to Blind Mice: Creating the First Pragmatics Understanding Benchmarks for Slovene*
Mojca Brglez and Spela Vintar

12:40–14:00 **Lunch**

14:00–14:40 **Keynote III: Maria Gavriilidou**

14:40–15:40 **Oral talks III**

14:40–15:00 *SdQuAD: A Large Benchmark Question Answering Dataset for Low-resource Sindhi Language*
Wazir Ali, Muhammad Rafay Shaikh, Nadia Ali and Amar Rehman

15:00–15:20 *LLMs as Assistants for Data Annotation: Addressing Disagreement and Supporting Expert Processes*
Mark Andrade, Bláithín Heffernan, Abigail Walsh and Sheila Castilho

15:20–15:40 *Annotation Quality in Aspect-Based Sentiment Analysis: A Case Study Comparing Experts, Students, Crowdworkers, and Large Language Models*
Niklas Donhauser, Jakob Fehle, Nils Constantin Hellwig, Markus Weinberger, Udo Kruschwitz and Christian Wolff

15:40–16:00 **Lightning poster presentations**

16:00–17:00 **Poster session**

Cross-Lingual Mathematical Reasoning in LLMs: Evaluating Performance on Icelandic vs. English Problems
Hafsteinn Einarsson

Struct2Unstruct: Creating Tender NER Datasets from Structured Procurement Records using Large Language Models
Asim Abbas, Mark Lee, Niloofer Shanavas, Venelin Kovatchev and Mubashir Ali

Link Prediction for Event Logs in the Process Industry
Anastasia Zhukova, Thomas Walton, Christian E. Lobmüller and Bela Gipp

MultiZebraLogic: A Multilingual Logical Reasoning Benchmark
Sofie Bruun and Dan Saattrup Smart

Progressing beyond Art Masterpieces or Touristic Clichés: how to assess your LLMs for cultural alignment?
António Branco, João Ricardo Silva, Nuno Marques, Luis M. S. Gomes, Ricardo Campos, Raquel Sequeira, Sara Nerea, Rodrigo Silva, Miguel Marques, Rodrigo Duarte, Artur Putyato, Diogo Folques and Tiago Valente

Evaluating Large Language Model-based Natural Language Generation for Modular Dialog systems
Vincent Emmerling, Christoph Kowalski, Amelie Sophie Robrecht-Hilbig and Stefan Kopp

JobResQA: Semi-Automatic Multilingual Benchmark Creation for LLM Machine Reading Comprehension on Résumés and Job Descriptions
Casimiro Pio Carrino, Paula Estrella, Rabih Zbib, Carlos Escolano and Jose A. R. Fonollosa

Beyond English and Evasion: A Human-Annotated Multi-Domain Benchmark for High-Stakes LLM Safety Evaluation in Chinese
Wajdi Zaghouani, Kholoud Khalil Aldous and Yicheng Gao

A multilingual hallucination benchmark
Freja Thoresen and Dan Saattrup Smart

Exploring the similarities and differences between VLM-driven and traditional OCR for Historical Swedish Data
Martin Johansson, Selma Waginder and Dana Dannélls

17:00–17:55 **Panel discussion**

17:55–18:00 **Closing**