



LREC 2026

**Learning Non-Literal Expressions with Small Data @
LREC 2026**

Workshop Proceedings

Editors
Markus Egg and Valia Kordoni

11 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 978-2-493814-80-7

Preface

Non-Literal Expressions (NLEs) in natural language are a reflection of fundamental cognitive processes such as analogical reasoning and categorisation, and are deeply rooted in everyday communication. NLEs understanding is therefore an essential task for language modeling. This task is especially challenging because it cannot be tackled by falling back on individual word meanings, but requires taking into account larger chunks of surrounding text or even contextual information. At the same time, it is important because the reliable processing of NLEs is relevant for optimizing downstream tasks like translation and summarization.

This workshop focuses on understanding of Non-Literal Expressions. While most of the earlier work on NLEs had been devoted to metaphor and metonymy, recent activities target other forms of NLEs as well, e.g., hyperbole (deliberate exaggeration), litotes (understatement), rhetorical questions, and irony. Humanly annotated corpora for NLEs have very recently started becoming available to the research community and may serve as the basis for data-driven approaches to NLEs processing, with the interrelated goals of first identifying and then interpreting such expressions. Such data is mostly of high linguistic quality, but still very limited in size. Thus, the workshop's focus is on adaptation of Language Models (LMs) and Deep Learning (DL) for processing of Non-Literal Expressions with limited high-quality data, since such constructs still pose big identification and processing challenges in natural language analysis tasks.

The workshop features contributions which focus on the use of techniques like self-training for leveraging unlabelled data, as well as in work that focuses on the incorporation of external linguistic resources and knowledge injection to enrich features, and also in research that describes work on utilisation of multitask learning with the aim to benefit from related tasks.

The workshop also wants to discuss alternative approaches which may elaborate on the use of pre-trained Language Models (LMs) as a foundation and the application of techniques like contrastive learning and clustering to identify challenging examples within the data, the ultimate aim of the workshop being to highlight the necessity of high-quality data, as well as cross-lingual datasets.

Organizing Committee

Markus Egg, Humboldt-Universität zu Berlin, Germany
Valia Kordoni, Humboldt-Universität zu Berlin, Germany

Programme Committee

Beata Beigman Klebanov, ETS, USA
Maria Berger, Ruhr-Universität Bochum, Germany
Yuri Bizzoni, Aarhus University, Denmark
Kenneth Church, VecML Inc., USA
Stefanie Dipper, Ruhr-Universität Bochum, Germany
Markus Egg, Humboldt-Universität zu Berlin, Germany
Anna Feldman, Montclair State University, USA
Debanjan Ghosh, Princeton, USA
Valia Kordoni, Humboldt-Universität zu Berlin, Germany
Emmy Liu, CMU, USA
Petya Osenova, Sofia University "St. Kl. Ohridski", Bulgaria
Sebastian Padó, IMS Stuttgart, Germany
Gudrun Reijnierse, Vrije Universiteit Amsterdam, The Netherlands
Sebastian Reimann, Ruhr-Universität Bochum, Germany
Adam Roussel, Ruhr-Universität Bochum, Germany
Tatjana Scheffler, Ruhr-Universität Bochum, Germany
Sabine Schulte im Walde, Universität Stuttgart
Vered Shwartz, The University of British Columbia, Canada
Caroline Sporleder, Georg-August-Universität Göttingen, Germany
Egon Stemle, EURAC, Italy

Table of Contents

<i>Challenges in Japanese Euphemism Classification: An Analysis of Pretrained Japanese and Multilingual Models</i>	
Noriko Takahashi, Whitney Poh, Libby Barak, Jing Peng and Anna Feldman	1
<i>Steering Pragmatic Interpretation in LLMs: A Diagnostic Evaluation of Few-Shot and Reasoning-Based Prompting for Indirect Speech Acts.</i>	
Massimiliano Orsini and Dominique Brunato	12
<i>Injecting Structured Lexicographic Knowledge into LLMs for Non-Literal Expression Disambiguation: A Controlled Study on Croatian</i>	
Slobodan Beliga, Ivana Filipović Petrović and Ana Meštrović.....	21
<i>Metaphor Identification in Spanish Oncological Discourse: The Role of Explicit Meaning in Low-Resource Settings</i>	
Lucia Pitarch, Jordi Bernad and Gemma Bel-Enguix.....	31
<i>Exploring Detection of Complex, Non-Literal Expressions of Cultural Motifs</i>	
Ibrahim H. Alyami and Mark A. Finlayson.....	40
<i>Artful Writing, Authentic Emotions: Distinguishing Human-Written from LLM-Generated Metaphors by Annotation and Classification</i>	
Michaela Regneri, Nooshin Aghajari and Thomas Kroedel.....	51
<i>Creation and Validation of a Monolingual Spanish NLI Dataset for Metaphor Interpretation via Model-in-the-Loop</i>	
Alec Sanchez-Montero, Gemma Bel-Enguix and SERGIO LUIS OJEDA TRUEBA	77
<i>A Hybrid Architecture for Metonymy Detection in Marathi</i>	
Pratibha Dongare	88
<i>Contextualising (Im)plausible Events Triggers Figurative Language</i>	
Annerose Eichel, Tonmoy Rakshit and Sabine Schulte im Walde	93
<i>A Novel Dataset and Three Ways to Approach Automatic Metaphor Detection in German Religious Online Forums</i>	
Sebastian Reimann and Tatjana Scheffler.....	106
<i>Decomposing Creativity: Two Small Datasets Combining Originality Ratings and Metaphor Annotations</i>	
Emilie Sitter, Sina Zarriß, Omar Momen and Berenike Herrmann.....	119

Workshop Program

Monday, May 11, 2026

9:00–13:00 Learning Non-Literal Expressions with Small Data

Room: 4

Chair: Valia Kordoni

9:00–9:10 *Introduction*

9:10–9:50 *Challenges in Japanese Euphemism Classification: An Analysis of Pre-trained Japanese and Multilingual Models*

Noriko Takahashi, Whitney Poh, Libby Barak, Jing Peng and Anna Feldman

9:50–10:10 *Steering Pragmatic Interpretation in LLMs: A Diagnostic Evaluation of Few-Shot and Reasoning-Based Prompting for Indirect Speech Acts.*

Massimiliano Orsini and Dominique Brunato

10:10–10:30 *Injecting Structured Lexicographic Knowledge into LLMs for Non-Literal Expression Disambiguation: A Controlled Study on Croatian*

Slobodan Beliga, Ivana Filipović Petrović and Ana Meštrović

10:30–11:00 *Coffee break*

11:00–11:40 *Poster session*

Metaphor Identification in Spanish Oncological Discourse: The Role of Explicit Meaning in Low-Resource Settings

Lucia Pitarch, Jordi Bernad and Gemma Bel-Enguix

Exploring Detection of Complex, Non-Literal Expressions of Cultural Motifs

Ibrahim H. Alyami and Mark A. Finlayson

Artful Writing, Authentic Emotions: Distinguishing Human-Written from LLM-Generated Metaphors by Annotation and Classification

Michaela Regneri, Nooshin Aghajari and Thomas Kroedel

Creation and Validation of a Monolingual Spanish NLI Dataset for Metaphor Interpretation via Model-in-the-Loop

Alec Sanchez-Montero, Gemma Bel-Enguix and SERGIO LUIS OJEDA TRUEBA

A Hybrid Architecture for Metonymy Detection in Marathi

Pratibha Dongare

Monday, May 11, 2026 (continued)

Contextualising (Im)plausible Events Triggers Figurative Language
Annerose Eichel, Tonmoy Rakshit and Sabine Schulte im Walde

11:40–12:00 *A Novel Dataset and Three Ways to Approach Automatic Metaphor Detection in German Religious Online Forums*
Sebastian Reimann and Tatjana Scheffler

12:00–12:20 *Decomposing Creativity: Two Small Datasets Combining Originality Ratings and Metaphor Annotations*
Emilie Sitter, Sina Zarriß, Omar Momen and Berenike Herrmann

12:20–13:00 *Unveiling Reasoning in Small Language Models: Insights into Literal and Non-Literal Understanding*
Debanjan Gosh