



LREC 2026

**10th Workshop on Linked Data in Linguistics
(LDL-2026) @ LREC 2026**

Workshop Proceedings

Editors

**John P. McCrae, Katerina Gkirtzou, Fahad Khan,
Patricia Martín Chozas, Sara Carvalho, and Erin
Canning**

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 978-2-493814-72-2

Preface

The Linked Data in Linguistics (LDL) workshop series has, over the years, established itself as an essential forum for the discussion of the latest trends in the application of Linked Data (LD) and Semantic Web technologies to linguistics and related fields such as digital lexicography, and the digital humanities. This year's edition of the workshop has a strong focus on the generation of LD datasets as well as the continued exploration of LD as a means of integrating heterogeneous data, both linguistic data as well as data deriving from related fields. In particular, LDL-2026 showcases a number of new projects and initiatives which aim to introduce new kinds of data and/or new kinds of resources into the linguistic linked data ecosystem, and which we therefore feel will be of great interest to the linguistic linked data community. For instance, Pellegrini et al's "Linked Open Data for West Nilotic Languages: The NILOMORPH project" presents a new project addressing the integration of complex, multilingual, and multimodal data on West Nilotic morphology via an LD approach. In "A Linguistic Ontology for Constructicography: the Research Constructicon and its Ontology Modules", instead, Winckel et al introduce the Research Constructicon (RCxn), a community-driven, ontology-based resource for representing constructions and their associated research processes. Meanwhile, McCrae, in "Open English NameNet: Extending English Wordnet with Names", presents a large-scale extension of WordNet with named entities from Wikidata, bridging lexical and encyclopedic knowledge. Billero's "Towards a Linguistic Linked Open Data Resource for Italian Cultural Heritage: The Lessico dei Beni Culturali Corpus" outlines efforts to represent the Lessico dei Beni Culturali as Linked Data, focusing on challenges in modeling historical and culturally-bound terminology. And in "Latin Represented Speech (LaReS): Linking LiLa and the DICES Database", Mambrini introduces a Linked Open Data (LOD) resource modeling represented speech in Latin literature through integration with the Linking Latin (LiLa) knowledge graph and established ontologies.

The workshop also showcases innovative uses of LD standards and technology as applied to linguistically relevant datasets. In "Towards the LinkEn Knowledge Base. A Neuro-Symbolic Approach to Build a Linked Data Hub for English Lemmas with Large Language Models", Augello and Passarotti introduce a lemma-centered knowledge base for English that combines OntoLex-Lemon modeling with a hybrid neuro-symbolic pipeline to support semi-automatic LD creation; Fiumanò et al's "Victim or Assailant? Exploring Narratives Through Knowledge Graph Queries" introduces DORIS, an ontology enabling cross-document framing analysis through knowledge graphs grounded in Frame Semantics; Chiarcos and Siewert's "Consolidating Syntactically Annotated Corpora with LLOD Technology. An Experiment in the Old Saxon Heliand" demonstrates how LLOD technologies can consolidate heterogeneous syntactic annotations of the Old Saxon Heliand using graph-based transformations; and Albertelli's "Modeling Topics as Linguistic Linked Open Data: a First Attempt Using BERTopic, OntoLex-Lemon and FrAC" explores the formalization of dynamically extracted topics from parliamentary discourse as Linked Data entities linked to linguistic and political metadata. Finally, McCrae et al's "Bridging the Gap Between Ontologies and Dictionaries: Requirements and Implementation of a New Core for OntoLex-Lemon" reports on a new core module for OntoLex-Lemon, addressing gaps in lexicographic modeling while ensuring compatibility with existing standards.

Taken together, the contributions to this year's workshop serve to highlight ongoing efforts to build interoperable language resources, to integrate heterogeneous linguistically relevant data, and to extend LD approaches to a greater variety of linguistic and cultural phenomena.

The Organizing Committee for LDL-2026

Organizing Committee

John P. McCrae, (University of Galway, Ireland)
Katerina Gkirtzou (Athena Research Center, Greece)
Fahad Khan (Consiglio Nazionale delle Ricerche and CLARIN-IT, Italy)
Patricia Martín Chozas (Universidad Politécnica de Madrid, Spain)
Sara Carvalho (University of Aveiro, Portugal)
Erin Canning (University of Oxford and Victoria and Albert Museum, United Kingdom)

Program Committee

Sina Ahmadi (University of Zurich, Austria)
Verginica Barbu Mititelu (Research Institute for Artificial Intelligence of the Romanian Academy, Romania)
Paul Buitelaar (Insight, Ireland)
Rute Costa (NOVA FCSH/NOVA CLUNL, Portugal)
Milan Dojchinovski (Czech Technical University, Czech Republic)
Francesca Frontini (CNR-ILC, Italy)
Frances Gillis Webber (University of Cape Town, South Africa)
Yoshihiko Hayashi (Waseda University, Japan)
Max Ionov (University of Zaragoza, Spain)
Alik Kirillovich (ex. Higher School of Economics, Russia)
Penny Labropoulou (Athena Research Center, Greece)
Chaya Liebeskind (Jerusalem College of Technology, Israel)
David Lindemann (University of the Basque Country, Spain)
Elena Montiel Ponsoda (Universidad Politécnica de Madrid, Spain)
Steven Moran (University of Neuchâtel, Switzerland)
Petya Osenova (IICT-BAS, Bulgaria)
Ana Ostroški Anić (Institute of Croatian Language and Linguistics, Croatia)
Antonio Pareja Lora (Universidad de Alcalá, Spain)
Giulia Pedonese (CNR-ILC, Italy)
Sigita Rackevičienė (Mykolas Romeris University, Lithuania)
Felix Sasaki (SAP, Germany)
Andrea Schalley (Karlstad University, Sweden)
Gilles Sérasset (University Grenoble Alpes, France)
Milena Slavcheva (IICT-BAS, Bulgaria)
Blerina Spahiu (Bicocca University, Italy)
Ranka Stanković (University of Belgrade, Serbia)
Armando Stellato (University of Rome, Italy)
Marieke van Erp (KNAW Humanities Cluster, The Netherlands)
Federica Vezzani (University of Padua, Italy)

Table of Contents

<i>Modeling Topics as Linguistic Linked Open Data: A First Attempt Using BERTopic, Ontolex-Lemon and FrAC</i>	
Lisa Sophie Albertelli	1
<i>Towards the LinkEn Knowledge Base. A Neuro-Symbolic Approach to Build a Linked Data Hub for English Lemmas with Large Language Models</i>	
Lorenzo Augello and Marco Passarotti	13
<i>Towards a Linguistic Linked Open Data Resource for Italian Cultural Heritage: The Lessico Dei Beni Culturali Corpus</i>	
Riccardo Billero	22
<i>Consolidating Syntactically Annotated Corpora with LLOD Technology. An Experiment in the Old Saxon Heliand</i>	
Christian Chiarcos and Janine Siewert	29
<i>Victim or Assailant? Exploring Narratives through Knowledge Graph Queries</i>	
Beatrice Fiumanò, Nicolas Lazzari, Simone Paolo Ponzetto and Valentina Presutti	40
<i>Latin Represented Speech (LaReS): Linking LiLa and the DICES Database</i>	
Francesco Mambrini	50
<i>Open English NameNet: Extending English Wordnet with Names</i>	
John P. McCrae	60
<i>Bridging the Gap between Ontologies and Dictionaries: Requirements and Implementation of a New Core for OntoLex-Lemon</i>	
John P. McCrae, Jorge Gracia, Fahad Khan and Philipp Cimiano	69
<i>Linked Open Data for West Nilotic Languages: The NILOMORPH Project</i>	
Matteo Pellegrini, Matthew Baerman and Oliver Bond	80
<i>A Linguistic Ontology for Constructicography: The Research Constructicon and its Ontology Modules</i>	
Elodie Winckel, Peter Uhrig and Stephanie Evert	87

Workshop Program

09:00–09:15	<i>Registration and Welcome</i>
09:15–10:00	<i>Invited Talk: "Making Knowledge Visible Again: Linguistic Linked Data in the Age of Large Language Models."</i> Johanna Monti
10:00–10:30	<i>Minute Madness and Poster Session</i>
10:30–11:00	<i>Coffee Break and Poster Session</i>
11:00–11:20	<i>Towards the LinkEn Knowledge Base. A Neuro-Symbolic Approach to Build a Linked Data Hub for English Lemmas with Large Language Models</i> Lorenzo Augello and Marco Passarotti
11:20–11:40	<i>Open English NameNet: Extending English Wordnet with Names</i> John P. McCrae
11:40–12:00	<i>Victim or Assailant? Exploring Narratives through Knowledge Graph Queries</i> Beatrice Fiumanò, Nicolas Lazzari, Simone Paolo Ponzetto and Valentina Presutti
12:00–12:20	<i>Modeling Topics as Linguistic Linked Open Data: A First Attempt Using BERTopic, Ontolex-Lemon and FrAC</i> Lisa Sophie Albertelli
12:20–12:40	<i>Consolidating Syntactically Annotated Corpora with LLOD Technology. An Experiment in the Old Saxon Heliand</i> Christian Chiarcos and Janine Siewert
12:40–13:00	<i>Bridging the Gap between Ontologies and Dictionaries: Requirements and Implementation of a New Core for OntoLex-Lemon</i> John P. McCrae, Jorge Gracia, Fahad Khan and Philipp Cimiano
13:00–13:05	<i>Closing remarks</i>

List of Posters

Linked Open Data for West Nilotic Languages: The NILOMORPH Project

Matteo Pellegrini, Matthew Baerman and Oliver Bond

A Linguistic Ontology for Constructicography: The Research Constructicon and its Ontology Modules

Elodie Winckel, Peter Uhrig and Stephanie Evert

Towards a Linguistic Linked Open Data Resource for Italian Cultural Heritage: The Lessico Dei Beni Culturali Corpus

Riccardo Billero

Latin Represented Speech (LaReS): Linking LiLa and the DICES Database

Francesco Mambrini