

LREC-COLING 2024

**The 9th Workshop on Linked Data in Linguistics:  
Resources, Applications, Best Practices  
(LDL-2024) @LREC-COLING-2024**

Workshop Proceedings

Editors

John P. McCrae, Christian Chiarcos, Katerina Gkirtzou,  
Maxim Ionov, Fahad Khan, Patricia Martín-Chozas and  
Elena Montiel-Ponsoda

25 May, 2024  
Torino, Italia

**The 9th Workshop on Linked Data in Linguistics:  
Resources, Applications, Best Practices  
(LDL-2024) @LREC-COLING-2024**

Copyright ELRA Language Resources Association (ELRA), 2024  
These proceedings are licensed under a Creative Commons  
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-38-8  
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association  
and the International Committee on Computational Linguistics

## Preface by the Program Chairs

The Linked Data in Linguistics (LDL) workshop series has established itself as the premier venue for discussing the application of Semantic Web technologies to the fields of linguistics, digital lexicography, and digital humanities (DH).

While recent years have witnessed a steady growth in the adoption of the technology in these areas, its uptake in other relevant domains, most notably in the case of natural language processing (NLP), continues to lag behind. This year, aside from embracing the full bandwidth of applications of LLOD technologies and the closely related area of knowledge graphs in linguistics, we welcome contributions addressing the application of LLOD technologies to NLP applications, as well as those dealing with emerging hot topics of future bridges between structured (linguistic) knowledge and neural methods.

In addition, this year's edition of the workshop will be a venue for in-depth discussions on community standards and best practices, and, above all, those related to the work of the W3C community groups OntoLex,<sup>1</sup> LD4LT<sup>2</sup> and BPMLOD.<sup>3</sup> To this end, it will include featured talks on the latest achievements, developments, and perspectives of these W3C Community Groups.

This year, we received a total of 19 submissions and accepted 8 of these papers for oral presentation and a further 7 of these papers for poster presentation. The papers covered a wide range of topics related to the application of linked data to linguistics. Several papers covered issues related to lexicography, especially those using the OntoLex-lemon module and it was applied to many languages including Latin, Bosnian, Croatian, Serbian and Proto-Indo-European. Further, we had papers examining lexicons for Portuguese borrowings in Asian languages and the Babylonian Talmud. In addition, several papers looked at extensions to the OntoLex model and challenges in lexicographic modelling including morphological description and diachronic analysis.

Several papers have also looked into the wider applications of linguistic linked data and have highlighted specific challenges in applications to the digital humanities. This includes works looking at challenges of museum cataloguing, scholarly information extraction and linking challenges in the humanities. Further, papers looked at the challenge of interoperability around lexicons through new platforms and online services. Finally, a new challenge was the use of linked data to connect lexicons with corpora and several papers tackled this challenge, by proposing new formats for the representation and linking of corpora as well as by extensions to the OntoLex-lemon model to enable such linking.

In addition to this rich array of papers, the feature talks for the three W3C community groups will be given by Penny Labrapoulou on the progress of the LD4LT group and Katerina Gkirtzou on the BPMLOD group. The OntoLex group meeting will focus on the developments of the model and will be led by Christian Chiarcos.

Overall, this is an exciting programme, reflecting the diversity of the research area and the many exciting research directions for linked data and its application to linguistic questions.

This workshop is organised in the scope of COST Action CA18209 NexusLinguarum,<sup>4</sup> supported by COST (European Cooperation in Science and Technology).

LDL 2024 Organisation Committee

---

<sup>1</sup>Ontology-Lexica Community Group, <https://www.w3.org/community/ontolex/>

<sup>2</sup>Linked Data in Language Technology Community Group, <https://www.w3.org/community/ld4lt/>

<sup>3</sup>Best Practices in Multilingual Linked Open Data, <https://www.w3.org/community/bpmlod/>

<sup>4</sup><https://nexuslinguarum.eu/>

## Organising Committee

Christian Chiarcos (University of Augsburg, Germany)  
Katerina Gkirtzou (Athena Research Center, Greece)  
Maxim Ionov (University of Cologne, Germany)  
Fahad Khan (Consiglio Nazionale delle Ricerche, Italy)  
John P. McCrae (University of Galway, Ireland)  
Elena Montiel Ponsoda (Universidad Politécnica de Madrid, Spain)  
Patricia Martín Chozas (Universidad Politécnica de Madrid, Spain)

## Program Committee

Sina Ahmadi (George Mason University, USA)  
Verginica Barbu Mititelu (Research Institute for Artificial Intelligence of the Romanian Academy, Romania)  
Paul Buitelaar (Insight, Ireland)  
Sara Carvalho (University of Aveiro, Portugal)  
Rute Costa (NOVA FCSH/NOVA CLUNL, Portugal)  
Milan Dojchinovski (Czech Technical University, Czech Republic)  
Agata Filipowska (Uniwersytet Ekonomiczny w Poznaniu, Poland)  
Francesca Frontini (CNR-ILC, Italy)  
Frances Gillis Webber (University of Cape Town, South Africa)  
Voula Giouli (Athena Research Center, Greece)  
Dagmar Gromann (University of Vienna, Austria)  
Yoshihiko Hayashi (Waseda University, Japan)  
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)  
Francesco Mambrini (Università Cattolica del Sacro Cuore, Italy)  
Monica Monachini (CNR-ILC, Italy)  
Steven Moran (University of Neuchâtel, Switzerland)  
Diego Moussallem (Paderborn University, Germany)  
Roberto Navigli ("La Sapienza" Università di Roma, Italy)  
Petya Osenova (IICT-BAS, Bulgaria)  
Ana Ostroški Anić (Institute of Croatian Language and Linguistics, Croatia)  
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)  
Federica Vezzani (University of Padua, Italy)

## Table of Contents

<i>LLODIA: A Linguistic Linked Open Data Model for Diachronic Analysis</i> Florentina Armaselu, Chaya Liebeskind, Paola Marongiu, Barbara McGillivray, Giedre Valunaite Oleskeviciene, Elena-Simona Apostol, Ciprian-Octavian Truica and Daniela Gifu . . . . .	1
<i>Cross-Lingual Ontology Matching using Structural and Semantic Similarity</i> Shubhanker Banerjee, Bharathi Raja Chakravarthi and John Philip McCrae . . . . .	11
<i>Querying the Lexicon der indogermanischen Verben in the LiLa Knowledge Base: Two Use Cases</i> Valeria Irene Boano, Marco Passarotti and Riccardo Ginevra . . . . .	22
<i>Defining an Ontology for Museum Critical Cataloguing Terminology Guidelines</i> Erin Canning . . . . .	32
<i>The MOLOR Lemma Bank: a New LLOD Resource for Old Irish</i> Theodorus Fransen, Cormac Anderson, Sacha Beniamine and Marco Passarotti . . . . .	37
<i>CHAMUÇA: Towards a Linked Data Language Resource of Portuguese Borrowings in Asian Languages</i> Fahad Khan, Ana Salgado, Isuri Anuradha, Rute Costa, Chamila Liyanage, John P. McCrae, Atul Kumar Ojha, Priya Rani and Francesca Frontini . . . . .	44
<i>LODinG: Linked Open Data in the Humanities</i> Jacek Kudera, Claudia Bamberg, Thomas Burch, Folke Gernert, Maria Hinzmann, Susanne Kabatnik, Claudine Moulin, Benjamin Raue, Achim Rettinger, Jörg Röpke, Ralf Schenkel, Kristin Shi-Kupfer, Doris Schirra, Christof Schöch and Joëlle Weis . . . . .	49
<i>DigitAnt: a platform for creating, linking and exploiting LOD lexica with heterogeneous resources</i> Michele Mallia, Michela Bandini, Andrea Bellandi, Francesca Murano, Silvia Piccini, Luca Rigobianco, Alessandro Tommasi, Cesare Zavattari, Mariarosaria Zinzi and Valeria Quochi . . . . .	55
<i>Teanga Data Model for Linked Corpora</i> John P. McCrae, Priya Rani, Adrian Doyle and bernardo stearns . . . . .	66
<i>The Services of the LiLa Knowledge Base of Interoperable Linguistic Resources for Latin</i> Marco Passarotti, Francesco Mambrini and Giovanni Moretti . . . . .	75
<i>An Annotated Dataset for Transformer-based Scholarly Information Extraction and Linguistic Linked Data Generation</i> Vayianos Pertsas, Marialena Kasapaki and Panos Constantopoulos . . . . .	84
<i>Linguistic LOD for Interoperable Morphological Description</i> Michael Rosner and Maxim Ionov . . . . .	94
<i>Modeling linking between text and lexicon with OntoLex-Lemon: a case study of computational terminology for the Babylonian Talmud</i> Flavia Sciolette . . . . .	103

*OntoLex Publication Made Easy: A Dataset of Verbal Aspectual Pairs for Bosnian, Croatian and Serbian*

Ranka Stanković, Maxim Ionov, Medina Bajtarević and Lorena Ninčević ..... 108

*Towards Semantic Interoperability: Parallel Corpora as Linked Data Incorporating Named Entity Linking*

Ranka Stanković, Milica Ikonić Nešić, Olja Perisic, Mihailo Škorić and Olivera Kitanović 115