

ACL-IJCNLP 2015

**The 53rd Annual Meeting of the  
Association for Computational Linguistics and the  
7th International Joint Conference on Natural Language  
Processing**

**Proceedings of the 4th Workshop on Linked Data in  
Linguistics: Resources and Applications (LDL-2015)**

July 31, 2015  
Beijing, China

©2015 The Association for Computational Linguistics  
and The Asian Federation of Natural Language Processing

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-941643-57-0

## Linked Data in Linguistics 2015. Introduction and Overview

After half a century of computational linguistics, quantitative typology, empirical, corpus-based study of language, and computational lexicography, researchers in computational linguistics, natural language processing (NLP) or information technology as well as in digital humanities are confronted with an immense wealth of linguistic resources; and these are not only growing in number, but also in their heterogeneity. Accordingly, the limited interoperability between linguistic resources has been recognized as a major obstacle for data use and re-use within and across discipline boundaries, and represents one of the prime motivations for adopting linked data in our field.

With the rise of the Semantic Web, new representation formalisms and novel technologies have become available, and different communities are becoming increasingly aware of the potential of these developments with respect to the challenges posed by the heterogeneity and multitude of linguistic resources available today. Many of these approaches follow the **Linked (Open) Data paradigm**.

The LDL workshop series and LDL-2015 are organized by the Open Linguistics Working Group to bring together researchers from various fields of linguistics, NLP, and IT to present and discuss principles, case studies, and best practices for representing, publishing and linking linguistic data collections, and aims to facilitate the exchange of technologies, ideas and resources across discipline boundaries, that (to a certain extent) find a material manifestation in the emerging LLOD cloud.

LDL-2015, collocated with ACL-IJCNLP 2015, the 53rd Annual Meeting of the Association of Computational Linguistics and the 7th Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing in July 2015 in Beijing, China, is the fourth workshop on Linked Data in Linguistics following LDL-2012 (March 2012 in Frankfurt am Main, Germany), LDL-2013 (Sep 2013 in Pisa, Italy), and LDL-2014 (May 2014, Reykjavik, Iceland), as well as more specialized events such as the workshops on Multilingual Linked Open Data for Enterprises (MLODE-2012: Sep 2012 in Leipzig, Germany, MLODE-2014: Sep 2014 in Leipzig, Germany), and Natural Language Processing and Linked Open Data (NLP&LOD-2013: Sep 2013 in Hissar, Bulgaria), and the theme session on Linked Data in Linguistic Typology (at the 10th Biennial Conference of the Association for Linguistic Typology, ALT-2013, Aug 2013 in Leipzig, Germany), as well as presentations, panels and informal meetings at various conferences.

LDL-2015 is organized by the *Open Linguistics Working Group* (OWLG) of the Open Knowledge Foundation and the **Ontology-Lexica Community (OntoLex) Group**<sup>1</sup>. Like LDL-2014, LDL-2015 is supported by the EU Projects LIDER and QTLeap: The project **Linked Data as an Enabler of Cross-Media and Multilingual Content Analytics for Enterprises Across Europe** (LIDER) aims to provide an ecosystem for the establishment of linguistic linked open data, as well as media resources metadata, for a free and open exploitation of such resources in multilingual, cross-media content analytics across Europe. The project **Quality Translation with Deep Language Engineering Approaches** (QTLeap) explores novel ways for attaining machine translation of higher quality that are opened by a new generation of increasingly sophisticated semantic datasets (including linked open data) and by recent advances in deep language processing.

For the 4th edition of the workshop on Linked Data in Linguistics, we invited contributions discussing the application of the Linked Open Data Paradigm to linguistic data in various fields of linguistics, natural language processing, knowledge management and information technology in order to present and discuss *principles, case studies, and best practices* for representing, publishing and linking mono- and multilingual linguistic and knowledge data collections, including corpora, grammars, dictionaries, wordnets, translation memories, domain specific ontologies etc.

---

<sup>1</sup><http://www.w3.org/community/ontolex>

In this regard, the Linked Data Paradigm provides an important step towards making linguistic data: i) easily and uniformly queryable, ii) interoperable and iii) sharable over the Web using open standards such as the HTTP protocol and the RDF data model. As a result of preceding LDL workshops and the activities of the communities involved, a considerable amount of linguistic linked open data resources has been established, so that our community is now increasingly aiming to shift the focus from resource creation to resource linking and further to the development of innovative applications of these resources in linguistics and NLP. For the current issue of LDL, we thus focus on *resources and applications*.

Accordingly, LDL-2015 provides a forum for researchers on natural language processing and semantic web technologies to present case studies and best practices on the exploitation of linguistic resources exposed on the Web for **natural language processing** applications, or other content-centered applications such as content analytics, knowledge extraction, etc. The availability of massive linked open knowledge resources raises the question how such data can be suitably employed to facilitate different NLP tasks and research questions. Following the tradition of earlier LDL workshops, we encouraged contributions to the Linguistic Linked Open Data (LLOD) cloud and research on this basis. In particular, this pertains to contributions that demonstrate an added value resulting from the combination of linked datasets and ontologies as a source for semantic information with linguistic resources published according to as linked data principles. Another important question to be addressed in the workshop is how natural language processing techniques can be employed to further facilitate the growth and enrichment of linguistic resources on the Web.

## **Invited Talk by Key-Sun Choi**

In addition to full and short papers/dataset descriptions, LDL-2015 will feature Key-Sun Choi as an invited speaker. Key-Sun Choi is professor of the Korea Advanced Institute of Science and Technology (KAIST), Korea, since 1988, where he had been Head of Computer Science Department (2006-2011) and recently founded the KAIST research group on Open Knowledge Convergence (since 2012). Key-Sun Choi has contributed to Department of Knowledge Service Engineering and Graduate School of Information Security in KAIST as a Joint Professor.

He founded and directed Korterm (Korea Terminology Research Center for Natural Language and Knowledge Engineering, 1998) and Bora (National Research Resource Bank for Language and Annotation, 2003). He had been an invited researcher in NEC C&C Lab of Japan (1987-1988), a visiting scholar of CSLI of Stanford University (1997), and an invited researcher of NHK Science & Technology Research Laboratories (2002). His areas of expertise are natural language processing, ontology and knowledge engineering, semantic web and linked data, and their infrastructure including text analytics. He served as President (2009-2010) of AFNLP (Asia Federation of Natural Language Processing), the President (2006) of Korean Cognitive Science Society, and the Secretary of ISO/TC37/SC4 for language resource management standards since 2002.

Recent key areas of his work are entity linking and predicate linking from text to DBpedia-like knowledge-bases and the enrichment of text, mainly for Korean. In his talk, he will focus on this line of research and address aspects of lexicalizing ontologies, mapping local properties to ontologies, extracting base ontologies and enhancing multilingualism in DBpedia.

As organizers of LDL-2015, we are happy to welcome Key-Sun Choi as key note speaker to the workshop and look forward for fruitful discussions on the interface of Semantic Web and NLP in Beijing.



**Organizers:**

Christian Chiarcos, Goethe-Universität Frankfurt, Germany  
Philipp Cimiano, Universität Bielefeld, Germany  
Nancy Ide, Vassar College, NY  
John P. McCrae, Universität Bielefeld, Germany  
Petya Osenova, University of Sofia, Bulgaria

**Program Committee:**

Eneko Agirre, University of the Basque Country, Spain  
Guadalupe Aguado, Universidad Politécnica de Madrid, Spain  
Claire Bonial, University of Colorado at Boulder, USA  
Peter Bouda, Interdisciplinary Centre for Social and Language Documentation, Portugal  
Antonio Branco, University of Lisbon, Portugal  
Martin Brümmer, University of Leipzig, Germany  
Paul Buitelaar, INSIGHT, NUIG Galway, Ireland  
Steve Cassidy, Macquarie University, Australia  
Nicoletta Calzolari, ILC-CNR, Italy  
Thierry Declerck, DFKI, Germany  
Ernesto William De Luca, University of Applied Sciences Potsdam, Germany  
Gerard de Melo, University of California at Berkeley  
Judith Eckle-Kohler, Technische Universität Darmstadt, Germany  
Francesca Frontini, ILC-CNR, Italy  
Jeff Good, University at Buffalo  
Asunción Gómez Pérez, Universidad Politécnica de Madrid  
Jorge Gracia, Universidad Politécnica de Madrid, Spain  
Walther v. Hahn, University of Hamburg, Germany  
Yoshihiko Hayashi, Waseda University, Japan  
Fahad Khan, ILC-CNR, Italy  
Seiji Koide, National Institute of Informatics, Japan  
Bettina Klimek, Universität Leipzig, Germany  
Lutz Maicher, Universität Leipzig, Germany  
Elena Montiel-Ponsoda, Universidad Politécnica de Madrid, Spain  
Steven Moran, Universität Zürich, Switzerland/Ludwig Maximilian University, Germany  
Sebastian Nordhoff, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany  
Antonio Pareja-Lora, Universidad Complutense Madrid, Spain  
Maciej Piasecki, Wrocław University of Technology, Poland  
Francesca Quattri (Hong Kong Polytechnic University, Hong Kong  
Laurent Romary, INRIA, France  
Felix Sasaki, Deutsches Forschungszentrum für Künstliche Intelligenz, Germany  
Andrea Schalley, Griffith University, Australia  
Gilles Sérasset, Joseph Fourier University, France  
Kiril Simov, Bulgarian Academy of Sciences, Sofia, Bulgaria  
Milena Slavcheva, JRC-Brussels, Belgium  
Armando Stellato, University of Rome, Tor Vergata, Italy  
Marco Tadic, University of Zagreb, Croatia  
Marieke van Erp, VU University Amsterdam, The Netherlands

Daniel Vila, Universidad Politécnica de Madrid  
Cristina Vertan, University of Hamburg, Germany  
Menzo Windhouwer, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

**Invited Speaker:**

Key-Sun Choi, KAIST, Korea



## Table of Contents

<i>From DBpedia and WordNet hierarchies to LinkedIn and Twitter</i> Aonghus McGovern, Alexander O'Connor and Vincent Wade .....	1
<i>A Linked Data Model for Multimodal Sentiment and Emotion Analysis</i> J. Fernando Sánchez-Rada, Carlos A. Iglesias and Ronald Gil .....	11
<i>Seeing is Correcting: curating lexical resources using social interfaces</i> Livy Real, Fabricio Chalub, Valeria dePaiva, Claudia Freitas and Alexandre Rademaker .....	20
<i>Sar-graphs: A Linked Linguistic Knowledge Resource Connecting Facts with Language</i> Sebastian Krause, Leonhard Hennig, Aleksandra Gabryszak, Feiyu Xu and Hans Uszkoreit . . . .	30
<i>Reconciling Heterogeneous Descriptions of Language Resources</i> John Philip McCrae, Philipp Cimiano, Victor Rodriguez-Doncel, Daniel Vila Suero, Jorge Gracia, Luca Matteis, Roberto Navigli, Andrejs Abele, Gabriela Vulcu and Paul Buitelaar .....	39
<i>RDF Representation of Licenses for Language Resources</i> V́ctor Rodriguez-Doncel and Penny Labropoulou .....	49
<i>Linking Four Heterogeneous Language Resources as Linked Data</i> Benjamin Siemoneit, John Philip McCrae and Philipp Cimiano .....	59
<i>EVALution 1.0: an Evolving Semantic Dataset for Training and Evaluation of Distributional Semantic Models</i> Enrico Santus, Frances Yung, Alessandro Lenci and Chu-Ren Huang .....	64
<i>Linguistic Linked Data in Chinese: The Case of Chinese Wordnet</i> Chih-Yao LEE and Shu-Kai HSIEH .....	70





## Conference Program

- 9:00-9:30** Introduction
- 9:30-10:30** Invited Talk: “DBpedia and Multilingualism”  
Key-Sun Choi
- 10:30-11:00** Coffee break
- 11:00-11:30** *From DBpedia and WordNet hierarchies to LinkedIn and Twitter*  
Aonghus McGovern, Alexander O’Connor and Vincent Wade
- 11:30-12:00** *A Linked Data Model for Multimodal Sentiment and Emotion Analysis*  
J. Fernando Sánchez-Rada, Carlos A. Iglesias and Ronald Gil
- 12:00-12:30** *Seeing is Correcting: curating lexical resources using social interfaces*  
Livy Real, Fabricio Chalub, Valeria dePaiva, Claudia Freitas and Alexandre Rade-maker
- 12:30-14:00** Lunch
- 14:00-14:30** *Sar-graphs: A Linked Linguistic Knowledge Resource Connecting Facts with Language*  
Sebastian Krause, Leonhard Hennig, Aleksandra Gabryszak, Feiyu Xu and Hans Uszkoreit
- 14:30-15:00** *Reconciling Heterogeneous Descriptions of Language Resources*  
John Philip McCrae, Philipp Cimiano, Victor Rodriguez-Doncel, Daniel Vila Suero, Jorge Gracia, Luca Matteis, Roberto Navigli, Andrejs Abele, Gabriela Vulcu and Paul Buitelaar
- 15:00-15:30** *RDF Representation of Licenses for Language Resources*  
Victor Rodriguez-Doncel and Penny Labropoulou
- 15:30-16:00** Coffee break
- 16:00-16:20** *Linking Four Heterogeneous Language Resources as Linked Data*  
Benjamin Siemoneit, John Philip McCrae and Philipp Cimiano
- 16:20-16:40** *EVALution 1.0: an Evolving Semantic Dataset for Training and Evaluation of Distributional Semantic Models*  
Enrico Santus, Frances Yung, Alessandro Lenci and Chu-Ren Huang
- 16:40-17:00** *Linguistic Linked Data in Chinese: The Case of Chinese Wordnet*  
Chih-Yao LEE and Shu-Kai HSIEH
- 17:00-17:30** Coffee break