

# Preface

The explosion of information technology has led to a substantial growth in quantity, diversity and complexity of web-accessible linguistic data. These resources become even more useful when linked. This volume provides an overview over recent developments, use cases, applications and recommendations for the application of the linked data paradigm to represent, exploit, store, and connect different types of linguistic data collections.

Recent relevant developments include: (1) Language archives for language documentation, with audio, video, and text transcripts from hundreds of (endangered) languages. (2) Typological databases with typological and geographical data about languages from all parts of the globe. (3) Development, distribution and application of lexical-semantic resources in Natural Language Processing. (4) Multi-layer annotations and semantic annotation of corpora by corpus linguists and computational linguists, often accompanied by the interlinking of corpora with lexical-semantic resources.

The general trend of providing data online is accompanied by newly developing possibilities to interconnect linguistic data and metadata. This includes general-purpose knowledge bases such as the DBpedia (a machine-readable edition of the Wikipedia), but also repositories with specific linguistic information about languages, as well as about linguistic categories and phenomena.

It is the challenge of our time to store, interlink and exploit this wealth of data, e.g., by modeling different language resources as Linked Data. The contributions assembled in this volume illustrate the band-width of applications of the Linked Data paradigm for representative types of language resources, including lexical-semantic resources, annotated corpora, typological databases as well as terminology and metadata repositories. The book includes representative applications from different fields, ranging from theoretical linguistics (e.g., typology) over applied linguistics (e.g., language documentation, lexicography, and corpus linguistics) to computational linguistics, Natural Language Processing and information technology.

This volume accompanies the Workshop on Linked Data in Linguistics (LDL-2012), held March 7th-9th, 2012 in Frankfurt a. M., Germany, organized by the

Open Linguistics Working Group (OWLG) of the Open Knowledge Foundation (OKFN). It assembles contributions of the workshop participants and, beyond this, it summarizes initial steps in the formation of a Linked Open Data cloud of linguistic resources, the Linguistic Linked Open Data cloud (LLOD).

Heidelberg and Leipzig,  
December 2011

*Christian Chiarcos*  
*Sebastian Nordhoff*  
*Sebastian Hellmann*

Linked Data in Linguistics

Representing and Connecting Language Data and  
Language Metadata

Chiarcos, C.; Nordhoff, S.; Hellmann, S. (Eds.)

2012, XIV, 218 p., Hardcover

ISBN: 978-3-642-28248-5