

Ontology and the Lexicon

The relation between ontologies and language is at the forefront of both natural language processing (NLP) and knowledge engineering. Ontologies, as widely used models in semantic technologies, have much in common with the lexicon. A lexicon organizes words as a conventional inventory of concepts, while an ontology formalizes concepts and their logical relations. A shared lexicon is the prerequisite for knowledge-sharing through language, and a shared ontology is the prerequisite for knowledge-sharing through information technology. In building models of language, computational linguists must be able to map accurately the relations between words and the concepts that they can be linked to. This book focuses on the integration of lexical resources and semantic technologies. It will be of interest to researchers and graduate students in NLP, computational linguistics and knowledge engineering, as well as in semantics, psycholinguistics, lexicology and morphology/syntax.

CHU-REN HUANG is Chair Professor in the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University, and Research Fellow at the Institute of Linguistics, Academia Sinica in Taiwan.

NICOLETTA CALZOLARI is Director of Research in Istituto di Linguistica Computazionale at the CNR in Italy.

ALDO GANGEMI is senior researcher at the Institute of Cognitive Science and Technology of CNR (Italy), and coordinator of the Semantic Technology Lab.

ALESSANDRO LENCI is a researcher in the Department of Linguistics at the University of Pisa.

ALESSANDRO OLTRAMARI is a Research Fellow in the Laboratory for Applied Ontology, Institute of Cognitive Science and Technology at the Italian National Research Council.

LAURENT PRÉVOT is an Associate Professor at the Université de Provence.

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

Studies in Natural Language Processing

Series Editor:

Steven Bird, University of Melbourne

Editorial Board Members:

Chu-Ren Huang, The Hong Kong Polytechnic University and
Academia Sinica

Christopher Manning, Stanford University

Yuji Matsumoto, Nara Institute of Science and Technology

Maarten de Rijke, University of Amsterdam

Harold Somers, Dublin City University

Suzanne Stevenson, University of Toronto

This series offers widely accessible accounts of the state-of-the-art in natural language processing (NLP). Established on the foundations of formal language theory and statistical learning, NLP is burgeoning with the widespread use of large annotated corpora, rich models of linguistic structure, and rigorous evaluation methods. New multilingual and multimodal language technologies have been stimulated by the growth of the web and pervasive computing devices. The series strikes a balance between statistical versus symbolic methods; deep versus shallow processing; rationalism versus empiricism; and fundamental science versus engineering. Each volume sheds light on these pervasive themes, delving into theoretical foundations and current applications. The series is aimed at a broad audience who are directly or indirectly involved in natural language processing, from fields including corpus linguistics, psycholinguistics, information retrieval, machine learning, spoken language, human-computer interaction, robotics, language learning, ontologies, and databases.

Also in the series

Douglas E. Appelt, *Planning English Sentences*

Madeleine Bates and Ralph M. Weischedel (eds.), *Challenges in Natural Language Processing*

Steven Bird, *Computational Phonology*

Peter Bosch and Rob van der Sandt, *Focus*

Pierette Bouillon and Federica Busa (eds.), *Inheritance, Defaults and the Lexicon*

Cambridge University Press & Assessment

978-1-009-34247-6 — *Ontology and the Lexicon: A Natural Language Processing Perspective*

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

Ronald Cole, Joseph Mariani, Hans Uszkoreit, Giovanni Varile, Annie Zaenen, Antonio Zampolli, and Victor Zue (eds.), *Survey of the State of the Art in Human Language Technology*

David R. Dowty, Lauri Karttunen, and Arnold M. Zwicky (eds.), *Natural Language Parsing*

Ralph Grishman, *Computational Linguistics*

Graeme Hirst, *Semantic Interpretation and the Resolution of Ambiguity*

András Kornai, *Extended Finite State Models of Language*

Kathleen R. McKeown, *Text Generation*

Martha Stone Palmer, *Semantic Processing for Finite Domains*

Terry Patten, *Systemic Text Generation as Problem Solving*

Ehud Reiter and Robert Dale, *Building Natural Language Generation Systems*

Manny Rayner, David Carter, Pierette Bouillon, Vassilis Digalakis, and Matis Wiren (eds.), *The Spoken Language Translator*

Michael Rosner and Roderick Johnson (eds.), *Computational Lexical Semantics*

Richard Sproat, *A Computational Theory of Writing Systems*

George Anton Kiraz, *Computational Nonlinear Morphology*

Nicholas Asher and Alex Lascarides, *Logics of Conversation*

Margaret Masterman (edited by Yorick Wilks), *Language, Cohesion and Form*

Walter Daelemans and Antal van den Bosch, *Memory-based Language Processing*

Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prévot (eds.), *Ontology and the Lexicon: A Natural Language Processing Perspective*

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

Ontology and the Lexicon

A Natural Language Processing Perspective

Edited by

Chu-Ren Huang

Department of Chinese and Bilingual Studies

The Hong Kong Polytechnic University

Institute of Linguistics, Academia Sinica

Nicoletta Calzolari

Istituto di Linguistica Computazionale del CNR, Italy

Aldo Gangemi

Institute of Cognitive Science and Technology

Italian National Research Council

Alessandro Lenci

Department of Linguistics

Università degli Studi, Pisa

Alessandro Oltramari

Institute of Cognitive Science and Technology

Italian National Research Council

Laurent Prévot

Laboratoire Parole et Langage

Université de Provence, France



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)



CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781009342476

© Cambridge University Press & Assessment 2010

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

First published 2010

First paperback edition 2022

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-88659-8 Hardback

ISBN 978-1-009-34247-6 Paperback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables, and other factual information given in this work is correct at the time of first printing but Cambridge University Press & Assessment does not guarantee the accuracy of such information thereafter.

Contents

<i>Contributors</i>	<i>page</i> xi
<i>Preface</i>	xvii
Part I Fundamental aspects	1
1 Ontology and the lexicon: a multidisciplinary perspective	3
1.1 Situating ontologies and lexical resources	3
1.2 The content of ontologies	10
1.3 Theoretical framework for the ontologies/lexicons interface	14
1.4 From ontologies to the lexicon and back	21
1.5 Outline of chapters	23
2 Formal ontology as interlingua: the SUMO and WordNet linking project and global WordNet	25
2.1 WordNet	25
2.2 Principles of construction of formal ontologies and lexicons	29
2.3 Mappings	30
2.4 Interpreting language	32
2.5 Global WordNet	33
2.6 SUMO translation templates	35
3 Interfacing WordNet with DOLCE: towards OntoWordNet	36
3.1 Introduction	36
3.2 WordNet's preliminary analysis	37
3.3 The DOLCE upper ontology	39
3.4 Mapping WordNet into DOLCE	48
3.5 Conclusion	52
4 Reasoning over natural language text by means of FrameNet and ontologies	53
4.1 Introduction	53
4.2 An introduction to the FrameNet lexicon	54
4.3 Linking FrameNet to ontologies for reasoning	56
4.4 Formalizing FrameNet in OWL DL	57
4.5 Reasoning over FrameNet-annotated text	62
4.6 Linking FrameNet to SUMO	66
	vii

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

viii	Contents	
	4.7 Discussion	69
	4.8 Conclusion and outlook	70
5	Synergizing ontologies and the lexicon: a roadmap	72
	5.1 Formal mappings between ontologies	72
	5.2 Evaluation of ontolex resources	73
	5.3 Bridging different lexical models and resources	75
	5.4 Technological framework	77
Part II	Discovery and representation of conceptual systems	79
6	Experiments of ontology construction with Formal Concept Analysis	81
	6.1 Introduction	81
	6.2 Basic concepts and related work	82
	6.3 Dataset selection and design of experiments	86
	6.4 Evaluation and discussion	92
	6.5 Conclusion and future work	96
7	Ontology, lexicon, and fact repository as leveraged to interpret events of change	98
	7.1 Introduction	98
	7.2 A snapshot of OntoSem	100
	7.3 Motivation for pursuing deep analysis of events of change	101
	7.4 <i>Increase</i>	102
	7.5 Content divorced from its rendering	114
	7.6 NLP with reasoning and for reasoning	117
	7.7 Conclusion	118
8	Hantology: conceptual system discovery based on orthographic convention	122
	8.1 Introduction: hanzi and conventionalized conceptualization	122
	8.2 General framework	126
	8.3 Conceptualization and classification of the radicals system	128
	8.4 The ontology of a radical as a semantic symbol	132
	8.5 The architecture of Hantology	133
	8.6 OWL encoding of Hantology	137
	8.7 Summary	139
	8.8 Conclusion	142
9	What's in a schema?	144
	9.1 Introduction	144
	9.2 An ontology for cognitive linguistics	146
	9.3 The <i>c.DnS</i> ontology	148
	9.4 Schemata, mental spaces, and constructions	161
	9.5 An embodied semiotic metamodel	166
	9.6 Applying Semion to FrameNet and related resources	169
	9.7 Conclusion	181

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

Contents	ix
Part III Interfacing ontologies and lexical resources	183
10 Interfacing ontologies and lexical resources	185
10.1 Introduction	185
10.2 Classifying experiments in ontologies and lexical resources	185
10.3 Ontologies and their construction	188
10.4 How actual resources fit the classification	190
10.5 Two practical examples	194
10.6 Available tools for the ontology lexical resource interface	196
10.7 Conclusion	200
11 Sinica BOW (Bilingual Ontological WordNet): integration of bilingual WordNet and SUMO	201
11.1 Background and motivation	201
11.2 Resources and structure required in the BOW approach	202
11.3 Interfacing multiple resources: a lexicon-driven approach	204
11.4 Integration of multiple knowledge sources	207
11.5 Updating and future improvements	209
11.6 Conclusion	210
12 Ontology-based semantic lexicons: mapping between terms and object descriptions	212
12.1 Introduction	212
12.2 Why we need semantic lexicons	213
12.3 More semantics than we need	215
12.4 The semantics we need is in ontologies	218
12.5 Conclusion	223
13 Merging global and specialized linguistic ontologies	224
13.1 Introduction	224
13.2 Linguistic ontologies versus formal ontologies	226
13.3 Specialized linguistic ontologies	229
13.4 The plug-in approach	230
13.5 Experiments	236
13.6 Applications and extensions	237
13.7 Conclusion	238
Part IV Learning and using ontological knowledge	239
14 The life cycle of knowledge	241
14.1 Introduction	241
14.2 Using ontological knowledge in NLP	242
14.3 Creating ontological knowledge with NLP	249
14.4 Conclusion	256
15 The Omega ontology	258
15.1 Introduction	258
15.2 Constituents of Omega	258
15.3 Structure of Omega	260
15.4 Construction of Omega via merging	263

x	Contents	
	15.5 Omega’s auxiliary knowledge sources	264
	15.6 Applications	266
	15.7 Omega 5 and the OntoNotes project	267
	15.8 Discussion and future work	268
	15.9 Conclusion	269
16	Automatic acquisition of lexico-semantic knowledge for question answering	271
	16.1 Introduction	271
	16.2 Lexico-semantic knowledge for QA	272
	16.3 Related work	274
	16.4 Extracting semantically similar words	275
	16.5 Using automatically acquired role and function words	279
	16.6 Using automatically acquired categorized NEs	280
	16.7 Evaluation	283
	16.8 Conclusion and future work	286
17	Agricultural ontology construction and maintenance in Thai	288
	17.1 Introduction	288
	17.2 A framework of ontology construction and maintenance	290
	17.3 Ontology acquisition from texts	291
	17.4 Ontology acquisitions from a dictionary and a thesaurus	301
	17.5 Integration into an ontological tree	306
	17.6 Conclusion	307
	<i>References</i>	309
	<i>Index</i>	335

Contributors

The editors

CHU-REN HUANG, CHIEF EDITOR' is a research fellow at the Institute of Linguistics, Academia Sinica and a chair professor at the Department of Chinese and Bilingual Studies at the Hong Kong Polytechnic University. He is a member of ICCL, and past president of ROCLING and LST. He has played an important role in the development of Chinese corpus and computational linguistics since he received his PhD in linguistics from Cornell University in 1987. He currently serves on the editorial boards of Cambridge Studies in Natural Language Processing, *Journal of Chinese Linguistics*, *Computational Linguistics and Chinese Language Processing*, *Journal of Chinese Information Processing*, and *Language Resources and Evaluation*. He directed or co-directed the construction of Sinica Corpus, CKIP Lexicon, Sinica BOW, and Chinese WordNet. He was an organizer of OntoLex 2005, 2006, as well as a special session on upper ontology and NLP for NLP-KE 2003. His recent research foci include discovery of conventionalized ontology in written language and ontology-driven comparative studies of cross-lingual lexica.

NICOLETTA CALZOLARI graduated in Philosophy from the University of Bologna and is Director of Research at Istituto di Linguistica Computazionale of the CNR in Pisa, Italy. She received an honorary doctorate in philosophy from the University of Copenhagen in 2007. She has promoted internationally the fields of Language Resources and Standardization, and her research fields include human language technology, corpus linguistics, lexical semantics, knowledge acquisition, integration and representation. She has coordinated a large number of international and national projects, currently coordinating the FLReNet EC project. She is a member of ICCL, of the ACL Exec, vice-president of ELRA, chair of the Scientific Board of CLARIN, and member of many International Committees and Advisory Boards. She has been invited speaker, member of program committees, organizer for numerous international conferences and workshops, and general chair of the last three LRECs

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

xii Contributors

and of COLING-ACL-2006. She is co-editor-in-chief, with Nancy Ide, of the journal *Language Resources and Evaluation*, Springer.

ALDO GANGEMI graduated in philosophy from the University of Rome ‘La Sapienza’ in 1989 and then studied medical informatics, knowledge representation, and knowledge engineering with grants from the Italian National Research Council (CNR). He is currently senior researcher at the CNR Institute of Cognitive Sciences and Technology in Rome, and head of the Semantic Technology Lab. He also co-founded the Laboratory for Applied Ontology. His interests include ontology engineering, knowledge management, the Semantic Web, semiotics, and cognitive linguistics. His work on ontology engineering and conceptual modelling dates back to 1992, when he pioneered the application of ontologies to knowledge organization systems. His current research focuses on pattern-based ontology design, collaborative modelling, hybridation of NLP, lexical and semantic resources, and meta-models for heterogeneous knowledge integration. He has published more than 100 refereed articles in proceedings of international conferences, journals, and books, and has been involved in several projects funded by both the European Union and industrial companies (either as research partner or consultant). He is responsible at CNR for the EU projects IKS, NeOn and BONY, and has worked in the seminal ontology-related EU projects Galen, WonderWeb, OntoWeb, Metokis, and the Eureka project IKF. He has co-organized the first Formal Ontology and Information Systems Conference FOIS 1998, and the Linguistic Resources and Evaluation Conference, LREC 2006, and he is chairing the next European Knowledge Engineering and Knowledge Management Conference, EKAW 2008, as well as several workshops such as WORM at ODBASE2004 (focusing on regulatory ontologies), LEGONT at ICAIL2003 (focusing on legal ontologies), CORONT at EKAW2004 (focusing on core ontologies), and OPSW at ISWC2005 (focusing on ontology design patterns for the Semantic Web), the EON06: Ontology Evaluation for the Web Workshop at WWW2006, OntoLex: the ontology/lexicon interface at ISWC2007, and SWAP2008. He serves in the Programme Committees of main conferences in his field (KR, ISWC, WWW, FOIS, LREC) and gives invited academic and industrial talks, courses, and tutorials. He has acted as a reviewer for the EU-IST framework programmes and the Italian Ministries of Research and Industry.

ALESSANDRO LENCI is a researcher at the linguistic department of Pisa University and a member of the Institute of Computational Linguistics (ILC-CNR) also in Pisa. He is the author of numerous publications and has taught various courses concerning ontologies and the lexicon. His research interests concern computational linguistics, natural language processing, and cognitive sciences. More particularly he has actively contributed to

the following topics: cognitive and computational models of the lexicon, verb aspect, acquisition and representation of lexical knowledge, computational models of syntax, natural language learning, term extraction, and ontology learning.

ALESSANDRO OLTRAMARI received the Master degree in philosophy at the University of Padua in 2000 and PhD in Cognitive Science at the University of Trento in 2006. He has been a Research Fellow at LOA since 2000, where he mainly studies the methodological and applicative aspects of the integration between ontologies and computational lexicons. In this context, he used to be co-organizer of the main event in the field, the workshop ‘OntoLex’, in 2004 (Lisbon), 2005 (Jeju Island, Korea), and 2006 (Genoa). He has been visiting research associate at Princeton University (USA), guest of the Laboratory of Cognitive Science (WordNet group). He is author of about thirty scientific papers in the field of formal and applied ontology, computational linguistics, philosophy of mind, and cognitive semantics. His scientific interests include the foundational aspects of artificial intelligence, analysis of metaphor, affective computing, and human-machine interaction. Since January 2007 he has also collaborated with Cogito S.r.l. in the development of linguistic technologies.

LAURENT PRÉVOT obtained his PhD in computer science and formal linguistics at Toulouse Research Institute of Computer Science (IRIT). He has been a post-doctoral research fellow for several years, first at the Laboratory for Applied Ontology (ISTC-CNR), where he has sharpened his competence in knowledge representation and ontological analysis; then at the Institute of Linguistics of the Academia Sinica (Taipei) where he has contributed to research on the interface between ontologies and lexical resources; and finally at the CNRS within the Syntax and Semantics Research Unit (CLLE-ERSS) of Toulouse University. He is now an associate professor in linguistics at Provence University and a member of the laboratory ‘Parole et Langage’. His research interests also include discourse semantics as well as semantics and ontology of interaction and communication.

Other contributors

COLLIN F. BAKER, Project Manager, International Computer Science Institute, San Francisco, USA

STEPHEN BEALE, Research Assistant Professor, Department of Computer Science and Electrical Engineering University of Maryland Baltimore County, USA

Cambridge University Press & Assessment

978-1-009-34247-6 — *Ontology and the Lexicon: A Natural Language Processing Perspective*

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

xiv Contributors

STEFANO BORGIO, Researcher, Laboratory for Applied Ontology, Institute for Cognitive Sciences and Technology, Trento, Italy

GOSSE BOUMA, Associate Professor, Department of Information Science, University of Groningen, The Netherlands

PAUL BUITELAAR, Senior Research Fellow and Head, Unit for Natural Language Processing, Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland

RU-YNG CHANG, PhD student, Multimedia Human-Machine Communication Lab, Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan, Taiwan

YA-MIN CHOU, Assistant Professor, National Taipei University, Taipei, Taiwan

CHRISTIANE FELLBAUM, Senior Research Scientist, Department of Computer Science, Princeton University, USA

NICOLA GUARINO, Head, Laboratory for Applied Ontology, Institute of Cognitive Science and Technology, Trento, Italy

EDUARD HOVY, Director, Natural Language Group, Information Sciences Institute, University of Southern California, Los Angeles, USA

AURAWAN IMSOMBUT, Researcher, Department of Knowledge Management Dhurakij Pundit University, Bangkok, Thailand

ASANEE KAWTRAKUL, Associate Professor, Department of Computer Engineering, Kasetsart University, Bangkok, Thailand and Deputy Director, National Electronics and Computer Technology Center, Bangkok, Thailand

HSIANG-BIN LI, Research Assistant, Institute of Information Science, Academia Sinica, Taipei, Taiwan

SUJIAN LI, Associate Professor, Institute of Computational Linguistics, Peking University, Beijing, China

WENJIE LI, Assistant Professor, Department of Computing, The Hong Kong Polytechnic University, Hong Kong

QIN LU, Professor, Department of Computing, The Hong Kong Polytechnic University, Hong Kong

BERNARDO MAGNINI, Senior Researcher, HLT Research Unit, Trento, Italy

CLAUDIO MASOLO, Research Scientist, Laboratory for Applied Ontology, Institute for Cognitive Sciences and Technology, Trento, Italy

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

Contributors

xv

MARJORIE MCSHANE, Research Associate Professor, Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore county, USA

JORI MUR, PhD Student, Department of Information Science, University of Groningen, The Netherlands

SRINI NARAYANAN, Group Leader, Artificial Intelligence, International Computer Science Institute, University of California, Berkeley, USA

SERGEI NIRENBURG, Professor, Department of Computer Science and Electrical Engineering, University of Maryland Baltimore County, USA

ADAM PEASE, Principal Consultant and CEO, Articulate Software, San Francisco, USA

PATRICK PANTEL, Senior Scientist, Yahoo! Labs, Santa Clara, USA and Research Assistant Professor, Information Sciences Institute, University of Southern California, Los Angeles, USA

ANDREW PHILPOT, Research Scientist, Information Sciences Institute, University of Southern California, Los Angeles, USA

JAN SCHEFFCZYK, Solution Architect, E.ON IS GmbH, Hannover, Germany

MANUELA SPERANZA, Researcher HLT Research Unit, Trento, Italy

LONNEKE VAN DER PLAS, Post-Doctoral Researcher, Language Technology Laboratory, Department of Linguistics, University of Geneva, Switzerland

Preface

Why should I read this book?

The mapping between knowledge representation and natural language is fast becoming a focal point of both knowledge engineering (KE) and computational linguistics (CL). Ontologies have a special role to play in this interface. They are essential stepping stones (a) from natural language to knowledge representation and manipulation and (b) from formal theories of knowledge to their application in (natural language) processing. Moreover, the emergence of the Semantic Web initiative presents a unique opportunity to bring research results in this area to real-world applications, at the leading edge of human-language technology. An essential and perhaps foundational aspect of the mapping between knowledge representation and natural language is the interface between ontologies and lexical resources. On the one hand, their integration includes, but is not restricted to, the use of ontologies (a) as language-independent structures of multilingual computational lexicons and (b) as powerful tools for improving the performance of existing lexical resources on various natural language processing (NLP) tasks such as word-sense disambiguation. On the other hand, lexical resources constitute a formidable source of information for generating ontological knowledge both at foundational and domain levels.

This current volume aims to be an essential general reference book on the interface between ontology and lexical resources. Given the fast developments in this new research direction, we introduce a general framework with a terminology to accommodate both ontological and lexical perspectives. However, to show its relevance and its direction of development, the focus of the book is on the current applications of such a framework. The chapters present an up-to-date overview of contributions to the field, from both CL and KE communities. Hence it constitutes a self-contained reference from introduction to the domain to the latest applications and tools developed.

Intended audience

We hope this book presents a productive dialogue among the following communities: those from the CL research community who are interested in ontology and ontologies, those from Semantic Web and KE communities who pay special attention to natural language, and those from the linguistic community who want to know how lexical knowledge can be formalized and computed. As suggested by the title, this book not only aims to reduce the gaps among these research communities, but does so on a solid ground provided by a precise framework and definitions on all sides. The book should be accessible to Master and PhD students and has been used as material for such a course.

Issues and scope

The scope of the book can be summarized by the issues it covers. The value and relevance of this book will be determined by how many of these issues remain valid or central to the research areas in the future:

- What are ontologies and lexical resources?
- What are the benefits of considering them simultaneously?
- What kind of integration can be made between these knowledge systems?
- What are the benefits of such an integration (for their respective research domains)?
- What are the applications benefiting from this integration?

Some more precise questions are addressed:

- How can NLP treatments help build knowledge resources?
- How are lexical resources used in NLP applications?
- Can we improve current knowledge resources and how?
- What is the role of formal ontology in this context?

In addition to a general framework to address the above issues, specialists already familiar with the domain will find state-of-the-art exploratory application of recent theoretical development as well as complete development of known techniques on new domains and languages. The book introduces a new perspective with a focus on Asian languages, bringing new challenges to this new area.

How the book came about

The majority of papers in this book were from the emergent OntoLex workshop series, initiated in 2000 and held in 2002, 2004, 2005, 2006, 2007, and 2008. A

selection of contributors from the 2002, 2004, 2005, and 2006 workshops were invited to submit their papers, complementing other invited submissions. The papers were rigorously reviewed by reviewers invited by the Studies in Natural Language Processing editorial board, as well as by the editors of this volume. They are also cross-commented on by other authors. The editors worked in parallel on the introductory and the roadmap chapters.

Notation

In this book, the following conventions have been adopted for referring to the various WordNet-related objects:

- WordNet is a registered proper name referring to the Princeton WordNet.
- The term ‘wordnet’ is a common noun that refers to all lexical knowledge bases built following to some extent the original Princeton WordNet design.
- Many language wordnets, however, do use the template of ‘Language WordNet’, such as Italian WordNet and Chinese WordNet. This is not to be confused with the original WordNet.

As for notational issues:

- *Words* or *terms* are in italics
- CONCEPTS are in small capitals
- Synsets, { *synonymous set*, *synset*, ... }, are in bracketed italics. Note that the set provided does not have to be complete. For example, the synset { *dog*, *domestic dog*, *Canis familiaris* } can be simply referred to by { *dog* }. Also, sometimes the synset is given by a word indexed by a number, e.g. { *dog_2* }. In this later case, the synset corresponds to the second sense of the word *dog*.
- **Relations** are in boldface.

Acknowledgements

We would like to thank first and foremost the three editorial assistants, Li-wu Chen, Helen Kai-yun Chen, and Sophia Y. M. Lee, who spent numerous hours correcting and typesetting more than four versions of all the papers. This book could not have been completed without them. Steven Bird, the editor in chief of the Cambridge Studies in Natural Language Processing (SNLP) series, as well as several anonymous reviewers, also painstakingly read through and made comments on different versions of this book. This book improved greatly because of their contributions. We would also like to thank all of the contributors, who persevered through the long process. In particular, we would like to thank the few authors whose papers we were not able to include in the end because of other editorial considerations. Their contribution to research on

Cambridge University Press & Assessment

978-1-009-34247-6 — Ontology and the Lexicon: A Natural Language Processing Perspective

Edited by C. Huang, N. Calzolari, A. Gangemi, A. Lenci, A. Oltramari, L. Prevot

Frontmatter

[More Information](#)

xx Preface

the ontolex interface is by no way diminished by this omission. Colleagues at the Chinese WordNet group of Academia Sinica who helped us set up a website as well as process different versions of the chapters are Javan Ming-wei Hsu and I-Li Su. Editors who helped us check grammatical and stylistic errors are Professor Kathleen Ahrens, Maureen Kealy, and Colleen Lippert. Lastly, we thank the Cambridge University Press team in charge of the SNLP series, in particular Helen Barton, for their firm support and guidance in seeing this book through the publication process. Any remaining errors are, of course, ours.