Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Pre´vot (eds.): Ontology and the Lexicon: a natural language processing perspective. (Studies in Natural Language Processing.)
Cambridge University Press, Cambridge, 2010, xx+339 pp, ISBN 9780521886598, UK £60.00, US $105.00

Andrea C. Schalley
Griffith University, Brisbane, QLD, Australia
e-mail: a.schalley@griffith.edu.au

The relationship between ontologies and natural language lexicons is a hotly debated one. An ontology is a formalized system of concepts (potentially of a specific domain) and the relations these concepts entertain. A lexicon, on the other hand, is the language component that contains the conventionalized knowledge of natural language speakers about lexical items (mostly words, but also morphemes and idioms). Ontologies 'operate' on the conceptual level, lexicons on the linguistic level. Ontologies systematize and relate concepts, lexicons systematize and relate words and other lexical items. However, as semantic relations between lexical items reflect meaning relatedness and meaning is essentially conceptual, both notions appear to be very close to one another (and are often wrongly used interchangeably). The interplay of and mapping between ontologies and lexical resources is therefore a vital and challenging field of research, one which has gained additional momentum and importance through the Semantic Web enterprise.

The volume under review is a collection of 17 papers addressing the ontology- lexicon interface from a Natural Language Processing (NLP) perspective. It is divided into four parts: (I) fundamental aspects, (II) discovery and representation of conceptual systems, (III) interfacing ontologies and lexical resources, and (IV) learning and using ontological knowledge. The editors are leading scholars in the field and have made a number of important contributions to the field themselves, some of which are described in the volume (editors are involved in authoring nearly half of the papers of the volume). The volume is aimed at being ''an essential general reference book on the interface between ontology and lexical resources'' (p. xvii). This aim is achieved, due to the spread of approaches,

1

perspectives, and applications represented in the volume and due to the connecting and roadmap papers written by volume editors (Chapts. 1, 5, 10, and 14). The majority of the papers go back to the OntoLex workshop series and have been presented between 2002 and 2006 at those workshops. That is, more recent contributions are not included, even though, as the editors themselves state, there are ''fast developments in this new research direction'' (p. xvii). The volume displays a refreshing international view on the topic by containing a number of contributions on Asian languages. From an editing point of view, the book exhibits two rather unusual features: (1) the table of contents does not list the authors of the respective contributions (but their sub-structuring by indicating the section headings), and (2) all references are combined in one listing at the end of the book (pp. 309–334), which in some cases results in the same reference being listed twice. Because of (1), the following discussion of each chapter also lists its authors in full.

Chapter 1, ''Ontology and the lexicon: a multidisciplinary perspective'' (pp. 3–24, by Laurent Pre´vot, Chu-Ren Huang, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, and Alessandro Oltramari), constitutes the introductory chapter to both the volume as a whole and to Part I on fundamental aspects. The chapter presents the theoretical background on which the volume is founded. Coming from different angles, it discusses ontologies and lexicons/lexical resources, different notions of these as well as their interface, and it contextualises the contributions to the volume, in particular with a view to NLP. The authors endeavour to present a unified framework, a task that appears to be a challenging one, given that the terminology in this field of research is still unstable and developing. An editing inconsistency that should be noted is that the outline of the chapters (Sect. 1.5) lists five (not the actual four) chapters for Part II, six (not the actual four) chapters for Part III, and eight (not the actual four) chapters for Part IV, although it only comments on the actually included 17 contributions. Thus, seven papers seem to have not been included in the final version of the volume, and it is unclear why they have been excluded (''other editorial considerations'' are mentioned in the preface, p. xix).

Chapter 2, ''Formal ontology as interlingua: the SUMO and WordNet linking project and global WordNet'' (pp. 25–35, by Adam Pease and Christiane Fellbaum) reports on the mapping efforts between WordNet and SUMO (the Suggested Upper Merged Ontology), i.e. the mapping between the biggest and most-deployed electronical lexical resource and one of the prominent upper (i.e. containing very general concepts) formal ontologies. Both

WordNet and SUMO are introduced (the latter as part of the section on WordNet, which is an indication for the intransparent structure of the chapter), and their mapping is discussed using rather simple examples. More explicitness on the benefits and applicability of such a mapping would have been beneficial. Also, the relevance of the last sections could have been made clearer in order to improve the coherence of the chapter.

Chapter 3, ''Interfacing WordNet with DOLCE: towards OntoWordNet'' (pp. 36–52, by Aldo Gangemi, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari) is an excellent contribution discussing alignment work on WordNet's noun taxonomy in order to improve WordNet's overall ontological adequacy. This is well motivated: WordNet, although a lexical resource, is often used as an ontology. In mapping WordNet to yet another prominent ontology, DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), WordNet is re- engineered as a formal ontology. While the paper delivers a very good explanation of DOLCE itself, a few open questions remain in the mapping discussion, such as what has been done to WordNet elements that have been deemed as problematic, and, if WordNet elements are ''rejected'' (p. 48) and thus not included in the mapping, which consequence this has for applications.

Chapter 4, ''Reasoning over natural language text by means of FrameNet and ontologies'' (pp. 53–78, by Jan Scheffczyk, Collin F. Baker, and Srini Narayanan) is one of the few examples in the field in which WordNet is not used as the lexical resource for mapping, but FrameNet (based on Fillmore's frame semantics). FrameNet does not focus on relations such as hyponymy but on roles (so-called frame-elements) that are defined for each frame (where a frame is a set of concepts associated with an event, etc.). The chapter describes how FrameNet and FrameNet- annotated sentences have been formalized in OWL DL, a Web Ontology Language species that makes FrameNet accessible to Description Logic reasoners, and how FrameNet has been linked to SUMO, which makes 'world knowledge' available for reasoning over natural language text.

After introducing crucial resources and different approaches to combining formal ontologies and WordNet or FrameNet in Chapts. 2–4, Part I concludes with Chapt. 5, ''Synergizing ontologies and the lexicon: a roadmap'' (pp. 72–78, by Alessandro Oltramari, Aldo Gangemi, Chu-Ren Huang, Nicoletta Calzolari, Alessandro Lenci, and Laurent Pre´vot). The chapter, conceived as a roadmap, aims at providing perspectives for future research in the field. It mainly gives an overview of available technologies and initiatives in the field. It is sometimes hard to follow its line

of argument, though; more background argumentation would have been helpful.

Part II of the volume concentrates on the discovery and representation of conceptual systems. It shows a number of ways in which ontological knowledge can be detected and how it might be represented. In Chapt. 6, ''Experiments of ontology construction with Formal Concept Analysis'' (pp. 81–97, by SuJian Li, Qin Lu, and Wenjie Li), Formal Concept Analysis (FCA, where a mathematization of concepts and their hierarchy allows for the analysis of conceptual data) is used to derive an ontology for a set of pre-specified ontology concepts. That is, while the concepts are given, relations are automatically generated using FCA. Two kinds of data sources are explored for this, one manually constructed, the other a large-scale corpus, and their results are compared, giving rise to a solid methodological discussion.

In Chapt. 7, ''Ontology, lexicon, and fact repository as leveraged to interpret events of change'' (pp. 98–121, by Marjorie McShane, Sergei Nirenburg, and Stephen Beale), another major approach is presented: ''OntoSem is a text-processing environment that takes as input raw text and carries out its tokenization, morphological analysis, syntactic analysis, and semantic analysis to yield text- meaning representations (TMRs).'' (p. 100) Amongst other things, it relies on an ontology, a lexicon, and a fact repository for text analysis, and captures meaning using a language-independent metalanguage grounded in an ontology. In the paper under discussion, OntoSem is used to interpret 'events of change'; in particular, increase and its synonyms and hyponyms are considered as representatives of events of change. While the notion of 'event of change' is, from a purely linguistic point of view, way too narrow and restricted (it comprises change of property value, direction of change, and quantity of change; and all examples have preconditions and/or effects, on the basis of which 'meaning procedures' can be run), the contribution differs from others in that it demonstrates an approach that aims for rigorous semantic interpretation of language to support high-end applications. The authors rightly state that the quality and depth of knowledge offered by knowledge bases ''is, across the board, not sufficient to support truly sophisticated applications, and automatic processes launched on these resources […] only aggravate this problem'' (p. 119).

Chapter 8, ''Hantology: conceptual system discovery based on orthographic convention'' (pp. 122–143, by Ya-Min Chou and Chu-Ren Huang), presents a fascinating idea, namely to investigate and interpret the 3,000-year old

Chinese writing system as a body of conventionalized knowledge. Naturally, only an orthographic convention such as the Chinese writing system, in which lexical units are orthographically classified according to semantic classes, lends itself to such an approach. Hanzi (the Chinese characters) are interpreted as reflecting an underlying ontology. This 'Hantology' (the Hanzi Ontology), is studied in the chapter, mapped to SUMO, and the result is encoded in OWL (Web Ontology Language), which makes it accessible for the Semantic Web initiative. A very interesting aspect is also the link that is drawn from Hanzi basic elements, the radicals, to Pustejovsky's Generative Lexicon.

Chapter 9, ''What's in a schema?'' (pp. 144–182, by Aldo Gangemi), is the final chapter in Part II of the volume. It aims at abstracting from individual interfaces, lexical standards or specific transformation methods (cf. p. 145), by providing an appropriate metamodel ('semiotic fac¸ade') in between the more intuitive semantics of different lexica and 'formal semantics'. The metamodelling task is an important one, and the system suggested is convincing and aims at having psychological reality. The chapter integrates knowledge from different disciplines and does so well.

Chapter 10, ''Interfacing ontologies and lexical resources'' (pp. 185–200, by Laurent Pré´vot, Stefano Borgo, and Alessandro Oltramari) introduces the same- named Part III. It gives an overview of the domain, and compares ontologies built following different methodologies as well as analyses their combinations with lexical resources. A methodology classification for combining ontologies and lexical resources is proposed, and it is discussed how actual resources fit this classification. The classification, had it been introduced at the beginning of the book (e.g. in the introductory chapter), would have given the reader a better foundation for the study of the volume. Overall, the paper is rather short but very instructive (sections such as the one on tools could have easily been expanded).

Chapter 11, ''Sinica BOW (Bilingual Ontological WordNet): integration of bilingual WordNet and SUMO'' (pp. 201–211, by Chu-Ren Huang, Ru-Yng Chang, and Hsiang-bin Lee), showcases an integration of WordNet, a handcrafted English-Chinese translation equivalents database (a bilingual resource, called ECTED), and SUMO. Thereby, it presents a full-scale ontolex mapping for a new language without a previously constructed WordNet.

Chapter 12, ''Ontology-based semantic lexicons: mapping between terms and object descriptions'' (pp. 212–223, by Paul Buitelaar), argues for organizing a semantic lexicon on the basis of an ontology, and briefly introduces a lexical model (LingInfo). Such an approach would leave the

semantics to the ontology and hence take it out of the linguistic description (which then merely refers to semantics), in contrast to what Buitelaar sees as the traditional view on semantic lexicons (where senses are assigned to lexical items, with the set of senses generally being open- ended). This promising approach is close to some 'purely' linguistic enterprises, such as ontolinguistics ('ontology-based linguistic') as introduced by Schalley and Zaefferer, which views meaning as being constrained by an underlying ontology.

Chapter 13, ''Merging global and specialized linguistic ontologies'' (pp. 224–238, by Manuela Speranza and Bernardo Magnini), a very clearly written and well argued contribution, completes Part III. It discusses the merging of ontologies that have different degrees of granularity, and focuses on the integration of specialized linguistic ontologies with global ones. Given this focus, a strong precedence criterion can be established, which allows to manage both information overlapping and inheritance in the ontology integration procedure, using a 'plug-in approach' and a limited number of 'plug-relations'. Experiments and applications using this approach are reported on, however, more information on the experiment outcomes and issues arisen in the application of the plug-in approach would have rounded off the picture.

Chapter 14, ''The life cycle of knowledge'' (pp. 241–257, by Alessandro Lenci), presents an excellent introduction to Part IV of the volume, which focuses on the interaction between ontolexical resources and NLP. The contribution, which contains many leads for the interested reader to follow up on, demonstrates that there is a bidirectional interaction in the sense that NLP is used for acquiring ontological knowledge on the one hand, while it also accesses this knowledge in applications to comply with the needs of NLP tasks. ''Better understanding of how knowledge can be automatically carved out of texts can lead to ontolexical resources that are more 'attuned' to the way knowledge is expressed with natural language. In turn, this promises to lead to a better way of creating knowledge resources that can boost performance in NLP technology'' (p. 257).

Chapter 15, ''The Omega ontology'' (pp. 258–270, by Andrew Philpot, Eduard Hovy, and Patrick Pantel), presents a ''shallow, lexically oriented, term taxonomy'' (p. 258) named Omega, which contains no formal concept definitions and few relations between concepts. It merges WordNet and Mikrokosmos through an interesting merging strategy; i.e. it merges a lexically oriented network with a resource which is meant to provide a language-neutral repository of concepts in the world, in order to facilitate

the merging of lower models into a functional whole. Several auxiliary knowledge sources (FrameNet, automatically mined knowledge, etc.) have also been integrated. Applications of Omega, such as information integration and multilingual question-answering, are briefly touched on in the rather shallow paper, and future work is outlined.

Chapter 16, ''Automatic acquisition of lexico-semantic knowledge for question answering'' (pp. 271–287, by Lonneke van der Plas, Gosse Bouma, and Jori Mur), is devoted to the automatic acquisition of lexico-semantic knowledge for an open-domain question-answering system for Dutch. The paper specifically focuses on named entities such as persons and organisations, and the approach presented is hence rather restricted. It concludes with an evaluation of the reported acquisition and its benefits.

The last chapter of the volume, Chapt. 17 on ''Agricultural ontology construction and maintenance in Thai'' (pp. 288–308, by Asanee Kawtrakul and Aurawan Imsombut), is concerned with ontology construction and maintenance for the domain of agriculture, carried out on the basis of text and other linguistic resources such as a domain-specific dictionary and a thesaurus. The methods used in the presented approach are specifically targeted and developed for the domain in question and not easily transferable to other domains. The language studied is Thai, and in order to use WordNet, Thai expressions are translated into English first, which might be expected to introduce additional noise to the results (although this is not explicitly discussed).

The editors have to be congratulated on carrying out a painstaking editing task and bringing together an interesting collection of cross-connected contributions, covering both theoretical aspects and NLP applications using ontolex methods and resources. For the newcomer, the book might at times be difficult to understand, as a lot of background knowledge is required. As the first volume providing an overview of the ontolex interface (to my knowledge), it constitutes a significant contribution to the field.