

GWC 2016

Proceedings of the 8th Global Wordnet Conference

**Verginica Barbu Mititelu, Corina Forăscu, Christiane Fellbaum, Piek Vossen
(Eds.)**

27–30 Jan, 2016
Bucharest, Romania



**Global
WordNet
Association**

©2016 Global Wordnet Association

ISBN 978-973-0-20728-6

Foreword

This eighth meeting of the international Wordnet community coincides with the 15th anniversary of the Global WordNet Association and the 30th anniversary of the Princeton WordNet. We are delighted to welcome old and new colleagues from many countries and four continents who construct wordnets, ontologies and related tools, as well as colleagues who apply such resources in a wide range of Natural Language Applications or pursue research in lexical semantics.

The number of wordnets has risen to over 150 and includes – besides all the major world languages – many less-studied languages such as Albanian and Nepali. Wordnets have become a principal tool in computational linguistics and NLP, and *wordnet*, *SemCor* and *synset* have entered the language as common nouns. Coming together and sharing some of the results of our work is an important part of the larger collaborative effort to better understand both universal and particular properties of human languages.

Many people have donated their time and effort to make this meeting possible: the review committee, the local organizers and their helpers (Eric Curea, Maria Mitrofan, Elena Irimia), our sponsors (PIM, QATAR Airways, Oxford University Press), EasyChair and our host, the Romanian Academy. Above all, thanks go to you, the contributors, for traveling to Bucharest to present your work, listen and discuss.

Corina Forăscu, Alexandru Ioan Cuza University of Iași & RACAI

Verginica Mititelu, RACAI

Christiane Fellbaum, Princeton University

Piek Vossen, VU University Amsterdam

Jan 2016

Organisers:

Local Chairs Verginica Mititelu and Corina Forăscu

Program Chairs Christiane Fellbaum and Piek Vossen

Program Committee:

- Eneko Agirre, University of the Basque Country, Spain
- Eduard Barbu, Translated.net, Italy
- Francis Bond, Nanyang Technological University, Singapore
- Sonja Bosch, University of South Africa, South Africa
- Alexandru Ceaușu, Euroscript Luxembourg S.à r.l., Luxembourg
- Dan Cristea, Alexandru Ioan Cuza University of Iași, Romania
- Agata Cybulska, VU University Amsterdam / Oracle Corporation, the Netherlands Tsvetana Dimitrova, Institute for Bulgarian Language, Bulgaria
- Marieke van Erp, VU University Amsterdam, the Netherlands
- Christiane Fellbaum, Princeton University, USA
- Darja Fiser, University of Ljubljana, Slovenia
- Antske Fokkens, VU University Amsterdam, the Netherlands
- Corina Forăscu, Alexandru Ioan Cuza University of Iași & RACAI, Romania Ales Horak, Masaryk University, Czech Republic
- Florentina Hristea, University of Bucharest, Romania
- Shu-Kai Hsieh, Graduate Institute of Linguistics at National Taiwan University, Taiwan Radu Ion, Microsoft, Ireland
- Hitoshi Isahara, Toyohashi University of Technology, Japan
- Ruben Izquierdo Bevia, VU University Amsterdam, the Netherlands
- Kaarel Kaljurand, Nuance Communications, Austria
- Kyoko Kanzaki, Toyohashi University of Technology, Japan
- Svetla Koeva, Institute for Bulgarian Language, Bulgaria
- Cvetana Krstev, University of Belgrade, Serbia
- Margit Langemets, Institute of the Estonian Language, Estonia
- Bernardo Magnini, Fondazione Bruno Kessler, Italy
- Verginica Mititelu, RACAI, Romania
- Sanni Nimb, Society for Danish Language and Literature, Denmark Kemal Oflazer, Carnegie Mellon University, Qatar
- Heili Orav, University of Tartu, Estonia
- Karel Pala, Masaryk University, Czech Republic
- Adam Pease, IPsoft, USA
- Bolette Pedersen, University of Copenhagen, Denmark
- Ted Pedersen, University of Minnesota, USA
- Maciej Piasecki, Wroclaw University of Technology, Poland Alexandre Rademaker, IBM Research, FGV/EMAp Brazil

- German Rigau, University of the Basque Country, Spain
- Horacio Rodriguez, Universitat Politecnica de Catalunya, Spain Shikhar Kr. Sarma, Gauhati University, India
- Roxane Segers, VU University Amsterdam, the Netherlands Virach Sornlertlamvanich, SIIT, Thammasart University, Thailand Dan Ștefănescu, Vantage Labs, USA
- Dan Tufiș RACAI, Romania
- Gloria Vasquez, Lleida University, Spain
- Zygmunt Vetulani, Adam Mickiewicz University, Poland
- Piek Vossen, VU University Amsterdam, the Netherlands

Additional Reviewers:

- Anna Feltracco
- Filip Ilievski
- Vojtech Kovar
- Egoitz Laparra
- Simone Magnolini
- Zuzana Neverilova
- Adam Rambousek

Invited Speakers:

Erhard Hinrichs, University of Tübingen

Invited Talks

Erhard Hinrichs: The Awful German Language: How to cope with the Semantics of Nominal Compounds in GermaNet and in Natural Language Processing

The title for my presentation borrows from Mark Twain’s well-known 1880 essay “The Awful German Language”, where Twain cites pervasive nominal compounding in German as one of the pieces of evidence for the “awfulness” of the language. Two much cited examples of noun compounds that are included in the Duden dictionary of German are Kraftfahrzeughaftpflichtversicherung (‘motor car liability insurance’) and Donaudampfschiffahrtsgesellschaft (‘Danube steamboat shipping company’). Any dictionary of German, including the German wordnet GermaNet, has to offer an account of such compound words. Currently, GermaNet contains more than 55,000 nominal compounds. As the coverage of nouns in GermaNet is extended, new noun entries are almost always compounds.

In this talk I will present an account of how to model nominal compounds in GermaNet with particular focus on the semantic relations that hold between the constituents of a compound, e.g., the WHOLE-PART relation in the case of Roboterarm (‘robot arm’) or the LOCATION relation in the case of Berghütte (‘mountain hut’). This account, developed jointly with Reinhild Barkey, Corina Dima, Verena Henrich, Christina Hoppermann, and Heike Telljohann, borrows heavily from previous research on semantic relations in theoretical linguistics, psycholinguistics, and computational linguistics.

The second part of the talk will focus on using the semantic modelling of nominal compounds in a word net for the automatic classification of semantic relations for (novel) compound words. Here, I will present the results of recent collaborative work with Corina Dima and Daniil Sorokin, using machine learning techniques such as support vector machines as well as deep neural network classifiers and a variety of publicly available word-embeddings, which have been developed in the framework of distributional semantics.

Table of Contents

Adverbs in Sanskrit Wordnet	1
<i>Tanuja Ajotikar and Malhar Kulkarni</i>	
Word Sense Disambiguation in Monolingual Dictionaries for Building Russian WordNet	9
<i>Daniil Alexeyevsky and Anastasiya V. Temchenko</i>	
Playing Alias - efficiency for wordnet(s)	15
<i>Sven Aller, Heili Orav, Kadri Vare and Sirli Zupping</i>	
Detecting Most Frequent Sense using Word Embeddings and BabelNet	21
<i>Harpreet Singh Arora, Sudha Bhingardive and Pushpak Bhattacharyya</i>	
Problems and Procedures to Make Wordnet Data (Retro)Fit for a Multilingual Dictionary	26
<i>Martin Benjamin</i>	
Ancient Greek WordNet Meets the Dynamic Lexicon: the Example of the Fragments of the Greek Historians	34
<i>Monica Berti, Yuri Bizzoni, Federico Boschetti, Gregory R. Crane, Riccardo Del Gratta and Tariq Yousef</i>	
IndoWordNet::Similarity- Computing Semantic Similarity and Relatedness using IndoWordNet	39
<i>Sudha Bhingardive, Hanumant Redkar, Prateek Sappadla, Dharendra Singh and Pushpak Bhattacharyya</i>	
Multilingual Sense Intersection in a Parallel Corpus with Diverse Language Families	44
<i>Giulia Bonansinga and Francis Bond</i>	
CILI: the Collaborative Interlingual Index	50
<i>Francis Bond, Piek Vossen, John McCrae and Christiane Fellbaum</i>	
YARN: Spinning-in-Progress	58
<i>Pavel Braslavski, Dmitry Ustalov, Mikhail Mukhin and Yuri Kiselev</i>	
Word Substitution in Short Answer Extraction: A WordNet-based Approach	66
<i>Qingqing Cai, James Gung, Maochen Guan, Gerald Kurlandski and Adam Pease</i>	
An overview of Portuguese WordNets	74
<i>Valeria de Paiva, Livy Real, Hugo Gonçalo Oliveira, Alexandre Rademaker, Cláudia Freitas and Alberto Simões</i>	
Towards a WordNet based Classification of Actors in Folktales	83
<i>Thierry Declerck, Tyler Klement and Antonia Kostova</i>	
Extraction and description of multi-word lexical units in plWordNet 3.0	88
<i>Agnieszka Dziob and Michał Wendelberger</i>	
Establishing Morpho-semantic Relations in FarsNet (a focus on derived nouns)	93
<i>Nasim Fakoornia and Negar Davari Ardakani</i>	
Using WordNet to Build Lexical Sets for Italian Verbs	101
<i>Anna Feltracco, Lorenzo Gatti, Elisabetta Jezek, Bernardo Magnini and Simone Magnolini</i>	
A Taxonomic Classification of WordNet Polysemy Types	106
<i>Abed Alhakim Freihat, Fausto Giunchiglia and Biswanath Dutta</i>	
Some strategies for the improvement of a Spanish WordNet	115
<i>Matias Herrera, Javier Gonzalez, Luis Chiruzzo and Dina Wonsever</i>	

An Analysis of WordNet’s Coverage of Gender Identity Using Twitter and The National Transgender Discrimination Survey	123
<i>Amanda Hicks, Michael Rutherford, Christiane Fellbaum and Jiang Bian</i>	
Where Bears Have the Eyes of Currant: Towards a Mansi WordNet	131
<i>Csilla Horváth, Ágoston Nagy, Norbert Szilágyi and Veronika Vincze</i>	
WNSpell: a WordNet-Based Spell Corrector	136
<i>Bill Huang</i>	
Sophisticated Lexical Databases - Simplified Usage: Mobile Applications and Browser Plugins For Wordnets	144
<i>Diptesh Kanojia, Raj Dabre and Pushpak Bhattacharyya</i>	
A picture is worth a thousand words: Using OpenClipArt library for enriching IndoWordNet	150
<i>Diptesh Kanojia, Shehzaad Dhuliawala and Pushpak Bhattacharyya</i>	
Using Wordnet to Improve Reordering in Hierarchical Phrase-Based Statistical Machine Translation	155
<i>Arefeh Kazemi, Antonio Toral and Andy Way</i>	
Eliminating Fuzzy Duplicates in Crowdsourced Lexical Resources	162
<i>Yuri Kiselev, Dmitry Ustalov and Sergey Porshnev</i>	
Automatic Prediction of Morphosemantic Relations	169
<i>Svetla Koeva, Svetlozara Leseva, Ivelina Stoyanova, Tsvetana Dimitrova and Maria Todorova</i>	
Tuning Hierarchies in Princeton WordNet	178
<i>Ahti Lohk, Christiane Fellbaum and Leo Vohandu</i>	
Experiences of Lexicographers and Computer Scientists in Validating Estonian Wordnet with Test Patterns	185
<i>Ahti Lohk, Heili Orav, Kadri Vare and Leo Vohandu</i>	
African WordNet: A Viable Tool for Sense Discrimination in the Indigenous African Languages of South Africa	193
<i>Stanley Madonsela, Mampaka Lydia Mojapelo, Rose Masubelele and James Mafela</i>	
An empirically grounded expansion of the supersense inventory	199
<i>Hector Martinez Alonso, Anders Johannsen, Sanni Nimb, Sussi Olsen and Bolette Pedersen</i>	
Adverbs in plWordNet: Theory and Implementation	209
<i>Marek Maziarz, Stan Szpakowicz and Michal Kalinski</i>	
A Language-independent Model for Introducing a New Semantic Relation Between Adjectives and Nouns in a WordNet	218
<i>Miljana Mladenović, Jelena Mitrović and Cvetana Krstev</i>	
Identifying and Exploiting Definitions in Wordnet Bahasa	227
<i>David Moeljadi and Francis Bond</i>	
Semantics of body parts in African WordNet: a case of Northern Sotho	234
<i>Mampaka Lydia Mojapelo</i>	
WME: Sense, Polarity and Affinity based Concept Resource for Medical Events	243
<i>Anupam Mondal, Dipankar Das, Erik Cambria and Sivaji Bandyopadhyay</i>	
Mapping and Generating Classifiers using an Open Chinese Ontology	249

Luis Morgado Da Costa, Francis Bond and Helena Gao

IndoWordNet Conversion to Web Ontology Language (OWL)	257
<i>Apurva Nagvenkar, Jyoti Pawar and Pushpak Bhattacharyya</i>	
A Two-Phase Approach for Building Vietnamese WordNet	261
<i>Thai Phuong Nguyen, Van-Lam Pham, Hoang-An Nguyen, Huy-Hien Vu, Ngoc-Anh Tran and Thi-Thu-Ha Truong</i>	
Extending the WN-Toolkit: dealing with polysemous words in the dictionary-based strategy	267
<i>Antoni Oliver</i>	
A language-independent LESK based approach to Word Sense Disambiguation	275
<i>Tommaso Petrolito</i>	
plWordNet in Word Sense Disambiguation task	282
<i>Maciej Piasecki, Paweł Kędzia and Marlena Orlińska</i>	
plWordNet 3.0 – Almost There	292
<i>Maciej Piasecki, Stan Szpakowicz, Marek Maziarz and Ewa Rudnicka</i>	
Open Dutch WordNet	302
<i>Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen and Piek Vossen</i>	
Verifying Integrity Constraints of a RDF-based WordNet	311
<i>Alexandre Rademaker and Fabricio Chalub</i>	
DEBVisDic: Instant Wordnet Building	320
<i>Adam Rambousek and Ales Horak</i>	
Samāsa-Kartā: An Online Tool for Producing Compound Words using IndoWordNet	325
<i>Hanumant Redkar, Nilesh Joshi, Sandhya Singh, Irawati Kulkarni, Malhar Kulkarni and Pushpak Bhattacharyya</i>	
Arabic WordNet: New Content and New Applications	333
<i>Yasser Rezagui, Lahsen Abouenour, Fettoum Krieche, Karim Bouzoubaa and Paolo Rosso</i>	
Hydra for Web: A Browser for Easy Access to Wordnets	342
<i>Borislav Rizov and Tsvetana Dimitrova</i>	
Towards a methodology for filtering out gaps and mismatches across wordnets: the case of plWordNet and Princeton WordNet	347
<i>Ewa Rudnicka, Wojciech Witkowski and Łukasz Grabowski</i>	
Folktale similarity based on ontological abstraction	355
<i>Marijn Schraagen</i>	
The Predicate Matrix and the Event and Implied Situation Ontology: Making More of Events	364
<i>Roxane Segers, Egoitz Laparra, Marco Rospocher, Piek Vossen, German Rigau and Filip Ilievski</i>	
Semi-Automatic Mapping of WordNet to Basic Formal Ontology	373
<i>Selja Seppälä, Amanda Hicks and Alan Ruttenberg</i>	
Augmenting FarsNet with New Relations and Structures for verbs	382
<i>Mehrnoush Shamsfard and Yasaman Ghazanfari</i>	
High, Medium or Low? Detecting Intensity Variation Among polar synonyms in WordNet	389
<i>Raksha Sharma and Pushpak Bhattacharyya</i>	

The Role of the WordNet Relations in the Knowledge-based Word Sense Disambiguation Task . . .	396
<i>Kiril Simov, Alexander Popov and Petya Osenova</i>	
Detection of Compound Nouns and Light Verb Constructions using IndoWordNet	404
<i>Dhirendra Singh, Sudha Bhingardive and Pushpak Bhattacharyya</i>	
Mapping it differently: A solution to the linking challenges	411
<i>Meghna Singh, Rajita Shukla, Jaya Saraswati, Laxmi Kashyap, Diptesh Kanojia and Pushpak Bhattacharyya</i>	
WordNet-based similarity metrics for adjectives	419
<i>Emiel van Miltenburg</i>	
Toward a truly multilingual GlobalWordnet Grid	424
<i>Piek Vossen, Francis Bond and John McCrae</i>	
This Table is Different: A WordNet-Based Approach to Identifying References to Document Entities	432
<i>Shomir Wilson, Alan Black and Jon Oberlander</i>	
WordNet and beyond: the case of lexical access	441
<i>Michael Zock and Didier Schwab</i>	