

GWC 2018

Proceedings of the 9th Global Wordnet Conference

**Francis Bond, Takayuki Kuribayashi,
Christiane Fellbaum, Piek Vossen (Eds.)**

8–12 January, 2018
Nanyang Technological University (NTU)
Singapore



**Global
WordNet
Association**

©2018 Global Wordnet Association

ISBN 978-981-11-7087-4

Foreword

The Ninth Global Wordnet Conference was held at Nanyang Technological University, Singapore from 9–12th January 2018.

The program combined the main conference with a special day on wordnets and word-embeddings and finished with a half day workshop on technology enhanced learning (TEL). There were 4 invited talks, 41 full papers, 15 posters and 4 invited talks on TEL. Including the papers on embeddings, there were 15 rejections: the acceptance rate for full papers was 58% a sign of the consistently high quality of papers submitted to the conference. Copyrights for the papers reside with the original authors.

The invited papers were *One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction* by Ng Hwee Tou (National University of Singapore), *How are you two related? Corpus-based Learning of Lexical Semantic Relations* by Vered Shwartz (Bar-Ilan University), *Inducing Interpretable Word Senses for WSD and Enrichment of Lexical Resources* by Alexander Panchenko (University of Hamburg) and *Using a Grammar Implementation to Teach Writing Skills* by Dan Flickinger (Stanford). As well as many papers on distributional semantics, there were some on extending the coverage of existing wordnets, linking wordnets to new resources (especially in the medical domain), using wordnets for teaching and many other topics. There were papers from 24 different countries with every continent except Antarctica represented.

The conference and workshops were partially supported by the NTU Centre for Liberal Arts and Social Sciences (CLASS) and the Singapore MOE TRF *Grant Syntactic Well-Formedness Diagnosis and Error-Based Coaching in Computer Assisted Language Learning using Machine Translation Technology*. Support for students came from the Global Wordnet Association. We would like to thank the programme committee for their thoughtful and timely reviews.

The conference homepage is <http://compling.hss.ntu.edu.sg/events/2018-gwc/>.

Francis Bond, Nanyang Technological University

Takayuki Kuribayashi, Nanyang Technological University

Christiane Fellbaum, Princeton University

Piek Vossen, VU University Amsterdam

January 2018

Organizers

Local Chairs Francis Bond and Takayuki Kuribayashi

Program Chairs Christiane Fellbaum, Francis Bond and Piek Vossen

Workshop Chairs Christiane Fellbaum, Francis Bond and Erhard Hinrichs

Program Committee

Adam Pease	Articulate Software, San Jose, USA
Ales Horak	Masaryk University
Alexandre Rademaker	IBM Research Brazil and EMap/FGV
Antoni Oliver	Universitat Oberta de Catalunya
Bolette Pedersen	University of Copenhagen
Corina Forăscu	Alexandru Ioan Cuza University of Iași
Dan Cristea	Alexandru Ioan Cuza University of Iași
Darja Fiser	University of Ljubljana
Diptesh Kanojia	IIT Bombay
Emiel van Miltenburg	Vrije Universiteit Amsterdam
Eneko Agirre	University of the Basque Country
Erhard Hinrichs	University of Tübingen
Ewa Rudnicka	Wroclaw University of Technology
Gerard de Melo	Rutgers University
German Rigau	University of the Basque Country (EPV/EHU)
Haldur Oim	University of Tartu
Heili Orav	University of Tartu
Horacio Rodriguez	Universitat Politècnica de Catalunya
Isahara Hitoshi	Toyohashi University of Technology
Janos Csirik	University of Szeged
John P. Mccrae	National University of Ireland, Galway
Kadri Vider	University of Tartu
Kevin Scannell	Saint Louis University
Kyoko Kanzaki	Toyohashi University of Technology
Maciej Piasecki	Wrocław University of Technology
Magnini Bernardo	FBK-IRST
Marten Postma	VU University Amsterdam
Mikhail Khodak	Princeton University
Rada Mihalcea	University of North Texas / Oxford University
Roxane Segers	VU University Amsterdam
Shan Wang	The Education University of Hong Kong
Shu-Kai Hsieh	National Taiwan Normal University

Sonja Bosch
Ted Pedersen
Timothy Baldwin
Tomaž Erjavec
Verginica Mititelu

Department of African Languages, University of South Africa
University of Minnesota, Duluth
The University of Melbourne
Dept. of Knowledge Technologies, Jožef Stefan Institute
Romanian Academy Research Institute for Artificial Intelligence

Invited Speakers

- Ng Hwee Tou, National University of Singapore
- Vered Shwartz, Bar-Ilan University
- Alexander Panchenko, University of Hamburg
- Dan Flickinger, Stanford

Invited Talks

Ng Hwee Tou: One Million Sense-Tagged Instances for Word Sense Disambiguation and Induction

Supervised word sense disambiguation (WSD) systems have achieved the best performance when evaluated on standard benchmark datasets. However, the lack of large amounts of sense-tagged data poses a major hurdle to scaling up supervised WSD systems to disambiguate all words of English. In this talk, I will present a semi-automatic approach to extract and annotate a large sense-tagged corpus. This one-million-word sense-tagged corpus has been publicly released since 2015 and has been used by other researchers working on automated WSD. When trained on this one-million-word sense-tagged corpus, the open source IMS (It Makes Sense) WSD system created in my research group achieves good performance on standard WSD tasks and another word sense induction task.

Vered Shwartz: How are you two related? Corpus-based Learning of Lexical Semantic Relations

Recognizing lexical semantic relations between words is an essential component in semantic applications such as question answering and recognizing textual entailment. In order to overcome lexical variability, such systems traditionally relied heavily on lexical resources such as WordNet.

In the main part of the talk I will discuss our work on automatic detection of lexical semantic relations from free text. This task stems from the limited coverage of lexical resources, both in terms of missing lexical items (proper names, new words) and missing relations between existing items. Typical approaches to address this task are either distributional, i.e. based on the word embeddings of the two target words, or path-based (pattern-based) approach, based on the words co-occurrences in the corpus. I will present our integrated path-based and distributional method for recognizing lexical semantic relations, which is currently the state-of-the-art in this task.

In the second part, I will raise some questions about the interplay of WordNet and word embeddings: is external lexical knowledge obsolete in the deep learning era? And if it isn't, then how can lexical knowledge from WordNet and other resources be incorporated into neural models for semantic applications?

Alexander Panchenko: Inducing Interpretable Word Senses for WSD and Enrichment of Lexical Resources

In this talk, we will discuss induction of sparse and dense word sense representations using graph-based approaches and distributional models. Induced senses are represented by a vector, but also a set of hypernyms, images, and usage examples, derived in an unsupervised and knowledge-free manner, which ensure interpretability of the discovered senses by humans. We showcase the usage of the induced representations for the tasks of word sense disambiguation and enrichment of lexical resources, such as WordNet.

Dan Flickinger: Using a Grammar Implementation to Teach Writing Skills

This paper presents an approach to grammar checking, using a large-scale HPSG grammar of English. The system has been used in a Language Arts & Writing course for McGraw-Hill Education in U.S. classrooms for the past ten years. It has helped over 50,000 primary school students, mostly native English speakers. We have given feedback on over 10 million sentences. The feedback is generated using mal-rules that identify errors with high precision. We are currently looking at extending the system to non-native speakers' English.

Table of Contents

BanglaNet: Towards a WordNet for Bengali Language	1
<i>K.M. Tahsin Hassan Rahit, Khandaker Tabin Hasan, Md. Al- Amin and Zahiduddin Ahmed</i>	
WME 3.0: An Enhanced and Validated Lexicon of Medical Concepts	10
<i>Anupam Mondal, Dipankar Das, Erik Cambria and Sivaji Bandyopadhyay</i>	
Using Context to Improve the Spanish WordNet Translation	17
<i>Alfonso Methol, Guillermo López, Juan Álvarez, Luis Chiruzzo and Dina Wonsever</i>	
Towards Cross-checking WordNet and SUMO Using Meronymy	25
<i>Javier Álvez and German Rigau</i>	
Comparing Two Thesaurus Representations for Russian	34
<i>Natalia Loukachevitch, German Lashevich and Boris Dobrov</i>	
Towards Mapping Thesauri onto plWordNet	44
<i>Marek Maziarz and Maciej Piasecki</i>	
Investigating English Affixes and their Productivity with Princeton WordNet	53
<i>Verginica Mititelu</i>	
Mapping WordNet Instances to Wikipedia	61
<i>John P. McCrae</i>	
Mapping WordNet Concepts with CPA Ontology	69
<i>Svetla Koeva, Cvetana Dimitrova, Valentina Stefanova and Dimitar Hristov</i>	
Improving Wordnets for Under-Resourced Languages Using Machine Translation	77
<i>Bharathi Raja Chakravarthi, Mihael Arcan and John P. McCrae</i>	
Semantic Feature Structure Extraction From Documents Based on Extended Lexical Chains	87
<i>Terry Ruas and William Grosky</i>	
Toward a Semantic Concordancer	97
<i>Adam Pease and Andrew Cheung</i>	
Using OpenWordnet-PT for Question Answering on Legal Domain	105
<i>Pedro Delfino, Bruno Cuconato, Guilherme Paulino Passos, Gerson Zaverucha and Alexandre Rademaker</i>	
Implementation of the Verb Model in plWordNet 4.0	113
<i>Agnieszka Dziob and Maciej Piasecki</i>	
Public Apologies in India - Semantics, Sentiment and Emotion	123
<i>Sangeeta Shukla and Rajita Shukla</i>	
Derivational Relations in Arabic WordNet	136
<i>Mohamed Ali Batita and Mounir Zrigui</i>	
Extending Wordnet to Geological Times	145
<i>Henrique Muniz, Fabricio Chalub, Alexandre Rademaker and Valeria De Paiva</i>	
Towards Emotive Annotation in plWordNet 4.0	153
<i>Monika Zaśko-Zielińska and Maciej Piasecki</i>	
The Company They Keep: Extracting Japanese Neologisms Using Language Patterns	163

James Breen, Timothy Baldwin and Francis Bond

Lexical-semantic resources: yet powerful resources for automatic personality classification	172
<i>Xuan-Son Vu, Lucie Flekova, Lili Jiang and Iryna Gurevych</i>	
Towards a principled approach to sense clustering –a case study of wordnet and dictionary senses in Danish	182
<i>Bolette Pedersen, Manex Agirrezabal, Sanni Nimb, Ida Olsen and Sussi Olsen</i>	
WordnetLoom – a Multilingual Wordnet Editing System Focused on Graph-based Presentation . . .	190
<i>Tomasz Naskręt, Agnieszka Dziob, Maciej Piasecki, Chakaveh Saedi and António Branco</i>	
Translation Equivalence and Synonymy: Preserving the Synsets in Cross-lingual Wordnets	200
<i>Oi Yee Kwong</i>	
Lexical Perspective on Wordnet to Wordnet Mapping	209
<i>Ewa Rudnicka, Francis Bond, Łukasz Grabowski, Maciej Piasecki and Tadeusz Piotrowski</i>	
ReferenceNet: a semantic-pragmatic network for capturing reference relations.	219
<i>Piek Vossen, Filip Ilievski and Marten Postrma</i>	
Wordnet-based Evaluation of Large Distributional Models for Polish	229
<i>Maciej Piasecki, Gabriela Czachor, Arkadiusz Janz, Dominik Kaszewski and Paweł Kędzia</i>	
Distant Supervision for Relation Extraction with Multi-sense Word Embedding	239
<i>Sangha Nam, Kijong Han, Eun-Kyung Kim and Key-Sun Choi</i>	
Cross-Lingual and Supervised Learning Approach for Indonesian Word Sense Disambiguation Task	245
<i>Rahmad Mahendra, Heninggar Septiantri, Haryo Akbarianto Wibowo, Ruli Manurung and Mirna Adriani</i>	
Recognition of Hyponymy and Meronymy Relations in Word Embeddings for Polish	251
<i>Gabriela Czachor, Maciej Piasecki and Arkadiusz Janz</i>	
Simple Embedding-Based Word Sense Disambiguation	259
<i>Dieke Oele and Gertjan Van Noord</i>	
Semi-automatic WordNet Linking using Word Embeddings	266
<i>Kevin Patel, Diptesh Kanojia and Pushpak Bhattacharyya</i>	
Multilingual Wordnet sense Ranking using nearest context	272
<i>E Umamaheswari Vasanthakumar and Francis Bond</i>	
Grammatical Role Embeddings for Enhancements of Relation Density in the Princeton WordNet .	284
<i>Kiril Simov, Alexander Popov, Iliana Simova and Petya Osenova</i>	
An Iterative Approach for Unsupervised Most Frequent Sense Detection using WordNet and Word Embeddings	293
<i>Kevin Patel and Pushpak Bhattacharyya</i>	
Automatic Identification of Basic-Level Categories	298
<i>Chad Mills, Francis Bond and Gina-Anne Levow</i>	
African Wordnet: facilitating language learning in African languages	306
<i>Sonja Bosch and Marissa Griesel</i>	
Hindi Wordnet for Language Teaching: Experiences and Lessons Learnt	314
<i>Hanumant Redkar, Rajita Shukla, Sandhya Singh, Jaya Saraswati, Laxmi Kashyap, Diptesh Kanojia,</i>	

Preethi Jyothi, Malhar Kulkarni and Pushpak Bhattacharyya

An Experiment: Using Google Translate and Semantic Mirrors to Create Synsets with Many Lexical Units	324
<i>Ahti Lohk, Mati Tombak and Kadri Vare</i>	
Context-sensitive Sentiment Propagation in WordNet	329
<i>Jan Kocoń, Arkadiusz Janz and Maciej Piasecki</i>	
ELEXIS - a European infrastructure fostering cooperation and information exchange among lexicographical research communities	335
<i>Bolette Pedersen, John McCrae, Carole Tiberius and Simon Krek</i>	
Enhancing the Collaborative Interlingual Index for Digital Humanities: Cross-linguistic Analysis in the Domain of Theology	341
<i>Laura Slaughter, Wenjie Wang, Luis Morgado Da Costa and Francis Bond</i>	
Estonian Wordnet: Current State and Future Prospects	347
<i>Heili Orav, Kadri Vare and Sirli Zupping</i>	
Further expansion of the Croatian WordNet	352
<i>Krešimir Šojat, Matea Filko and Antoni Oliver</i>	
Linking WordNet to 3D Shapes	358
<i>Angel X Chang, Rishi Mago, Pranav Krishna, Manolis Savva and Christiane Fellbaum</i>	
Multisłownik: Linking plWordNet-based Lexical Data for Lexicography and Educational Purposes	364
<i>Maciej Ogrodniczuk, Joanna Bilińska, Zbigniew Bronk and Witold Kieraś</i>	
Putting Figures on Influences on Moroccan Darija from Arabic, French and Spanish using the WordNet	372
<i>Khalil Mrini and Francis Bond</i>	
pyiwn: A Python based API to access Indian Language WordNets	378
<i>Ritesh Panjwani, Diptesh Kanojia and Pushpak Bhattacharyya</i>	
Sinitic Wordnet: Laying the Groundwork with Chinese Varieties Written in Traditional Characters	384
<i>Chih-Yao Lee and Shu-Kai Hsieh</i>	
Synthesizing Audio for Hindi WordNet	388
<i>Diptesh Kanojia, Preethi Jyothi and Pushpak Bhattacharyya</i>	
Toward Constructing the National Cancer Institute Thesaurus Derived WordNet (ncitWN)	394
<i>Amanda Hicks, Selja Seppälä and Francis Bond</i>	
Towards a Crowd-Sourced WordNet for Colloquial English	401
<i>John P. McCrae, Ian Wood and Amanda Hicks</i>	
WordNet Troponymy and Extraction of "Manner-Result" Relations	407
<i>Aliaksandr Huminski and Hao Zhang</i>	
SardaNet: a Linguistic Resource for Sardinian Language	412
<i>Manuela Angioni, Franco Tuveri, Maurizio Viridis, Laura Lucia Lai and Micol Elisa Maltesi</i>	
Extraction of Verbal Synsets and Relations for FarsNet	420
<i>Fatemeh Khalghani and Mehrnoush Shamsfard</i>	