

GWC 2014 Tartu, Estonia

Proceedings of the Seventh Global Wordnet Conference

Tartu, Estonia, January 25-29, 2014

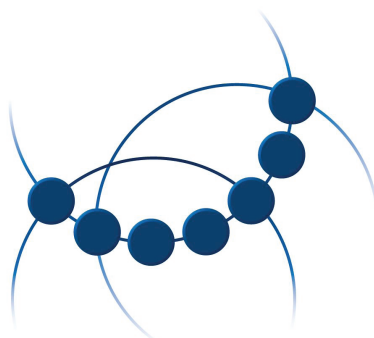
Editors: Heili Orav, Christiane Fellbaum, Piek Vossen



European Union
Regional Development Fund



Investing in your future



**CLARIN
ERIC**

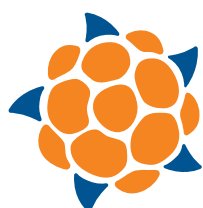
Common Language Resources and Technology Infrastructure



city of good thoughts



ESTONIAN COMPUTING



Center of
Estonian Language
Resources

Volume editors

Heili Orav
University of Tartu
e-mail: heili.orav@ut.ee

Christiane Fellbaum
Princeton University
e-mail: fellbaum@princeton.edu

Piek Vossen
VU University Amsterdam
e-mail: piek.vossen@vu.nl

ISBN 978-9949-32-492-7

ORGANIZATION

The seventh Global Wordnet Conference is organized by the University of Tartu, Institute of Computer Science in co-operation with the Global WordNet Association.

The conference homepage can be found at <http://gwc2014.ut.ee/>

PROGRAMME COMMITTEE

Eneko Agirre (University of the Basque Country), Francis Bond (Nanyang Technological University), Sonja Bosch (University of South Africa), Agata Cybulska (VU University Amsterdam), Christiane Fellbaum (Princeton University), Darja Fišer (University of Ljubljana), Yoshihiko Hayashi (Osaka University), Ales Horak (Masaryk University), Chu-Ren Huang (The Hong Kong Polytechnic University), Hitoshi Isahara (Toyohashi University of Technology), Kaarel Kaljurand (University of Zuerich), Kyoko Kanzaki (National Institute of Information and Communications Technology), Adam Kilgarriff (Lexical Computing Ltd), Kow Kuroda (National Institute of Information and Communications Technology), Margit Langemets (Institute of the Estonian Language), Haldur Õim (University of Tartu), Heili Orav (University of Tartu), Adam Pease (Articulate Software), Bolette Pedersen (University of Copenhagen), Ted Pedersen (University of Minnesota), Maciej Piasecki (Wroclaw University of Technology), German Rigau (IXA Group, UPV/EHU), Horacio Rodriguez (Universitat Politècnica de Catalunya), Virach Sornlertlamvanich (National Electronics and Computer Technology Center), Takenobu Tokunaga (Tokyo Institute of Technology), Gloria Vazquez (Universitat de Lleida), Zygmunt Vetulani (Adam Mickiewicz University), Kadri Vider (University of Tartu), Piek Vossen (VU University Amsterdam)

ORGANIZING COMMITTEE

Heili Orav (Chair)

Kairit Šor (Secretary)

Sven Aller (Homepage)

Sirli Parm, Kadri Vare, Katrin Alekand, Ingmar Jaska, Helen Türk, Eleri Aedma, Liisi Pool (Helpers)

Christiane Fellbaum, Piek Vossen (Co-organisers)

ADDITIONAL REVIEWERS

Kahusk, Neeme

Kubis, Marek

Marciniak, Jacek

Neverilova, Zuzana

Obrebski, Tomasz

Šmerk, Pavel

Preface

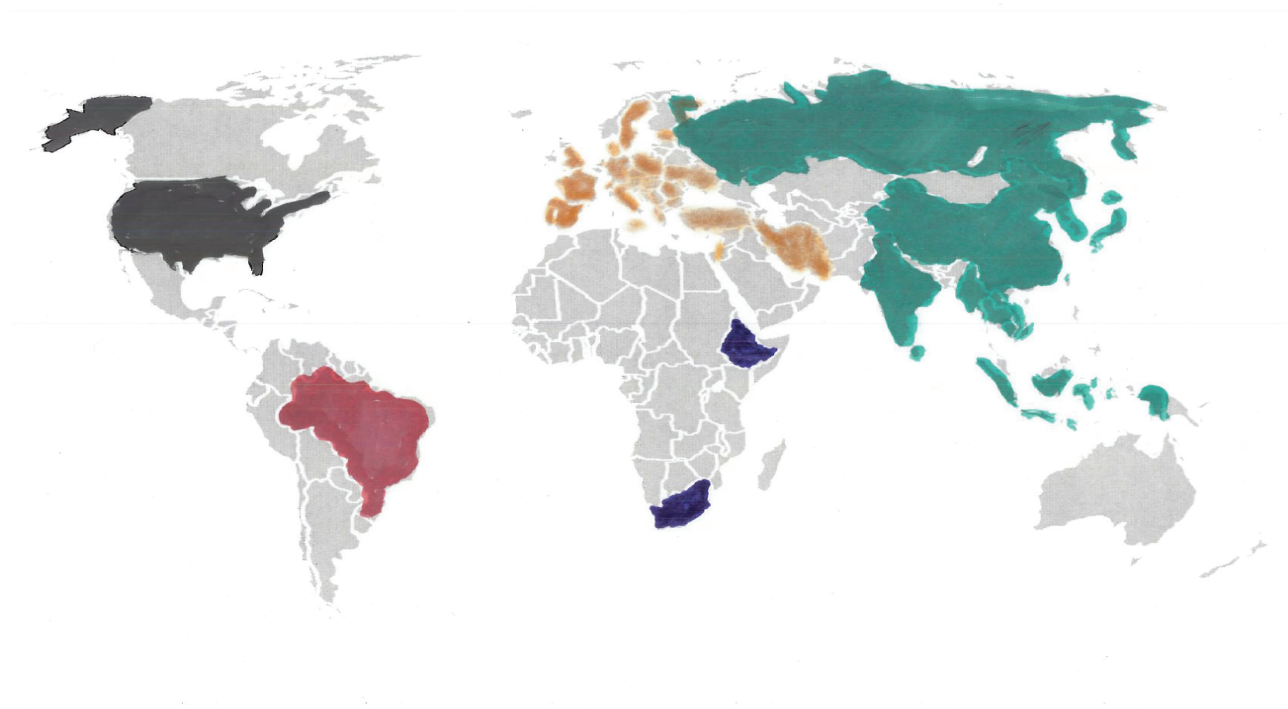
The seventh Global WordNet Conference includes presentations about new wordnets in languages like Amharic, Kurdish and Northern Sotho. The map shows the countries where wordnets are built in the local languages; if one colored in all the regions where these languages are spoken, most of the world would be covered!

Beyond the emergence of new lexical resources, the global wordnet endeavor has generated and facilitated research in linguistics, computational linguistics, psycholinguistics, ontology, lexicology, mathematics and a wide range of practical applications. The presentations in this volume reflect the manifold activities of our thriving global wordnet community.

We are grateful to the colleagues who reviewed submissions and provided constructive criticism as well as to the local organizers who performed uncountable large and small tasks. And we thank all of you present here for making this an exciting meeting.

Tartu, January 2014

Christiane Fellbaum, Piek Vossen, Heili Orav



Invited speaker: Alessandro Lenci

Will Distributional Semantics Ever Become Semantic?

Computational Linguistics Laboratory
Dept. of Philology, Literature, and Linguistics
University of Pisa (Italy)

`alessandro.lenci@ling.unipi.it`

Abstract

Distributional Semantics (DS) is a rich family of computational models that build semantic representations of lexical items from their statistical distribution in linguistic contexts. DS is currently experiencing an unprecedented fortune with a growing attention not only in computational linguistics, but also in cognitive science and theoretical linguistics. This is proved by the wide range of DS models that have appeared (e.g., vector spaces, Bayesian models, neural networks, etc.), but even more by the increased number of semantic tasks that these models have been applied to.

DS was born to address a specific issue, that is measuring the semantic similarity of lexical items to be used for thesaurus construction or synonym identification. The Distributional Hypothesis, the main theoretical foundation of DS, is in fact a statement about lexical semantic similarity, which is defined in terms of similarity of linguistic contexts. However, human semantic competence well exceeds the ability to judge lexical similarity. Polysemy, compositionality, inference, semantic creativity are only some of the main phenomena that must be part of the agenda of any full-fledged semantic theory. DS aims at becoming a general model for semantic representation and processing, and therefore it must be evaluated with respect to its ability to explain semantic facts like these. What is the current ability of DS to address these issues? To what extent semantic properties can be modeled in terms of distributional semantic similarity, or alternatively, can DS go beyond the mere notion of semantic similarity? What lies beyond its possibilities? Recently, DS has begun to address issues such as compositionality, polysemy, and semantic relations, but lots of questions remain open. The purpose of this talk is to explore the current boundaries of DS and the chances to enlarge them, in particular by finding new synergies with other types of semantic models.

Table of Contents

Towards Building KurdNet, the Kurdish WordNet.....	1
<i>Purya Aliabadi, Mohammad Sina Ahmadi, Shahin Salavati and Kyumars Sheykh Esmaili</i>	
WN-Toolkit: Automatic generation of WordNets following the expand model	7
<i>Antoni Oliver</i>	
Onto.PT: recent developments of a large public domain Portuguese wordnet	16
<i>Hugo Gonalo Oliveira and Paulo Gomes</i>	
Lexico-Semantic Annotation of <i>Skladnica</i> Treebank by means of P ₁ WN Lexical Units	23
<i>Elzbieta Hajnicz</i>	
WoNeF, an improved, expanded and evaluated automatic French translation of WordNet .	32
<i>Quentin Pradet, Gaël de Chalendar and Jeanne Baquénier-Desormeau</i>	
Bringing together over- and under- represented languages: Linking WordNet to the SIL Semantic Domains	40
<i>Muhammad Zulhelmy Bin Mohd Rosman, Frantisek Kratochvil and Francis Bond</i>	
Modeling Prefix and Particle Verbs in GermaNet	49
<i>Christina Hoppermann and Erhard Hinrichs</i>	
Developing and Maintaining a WordNet: Procedures and Tools	55
<i>Miljana Mladenović, Jelena Mitrović and Cvetana Krstev</i>	
Aligning Word Senses in GermaNet and the DWDS Dictionary of the German Language .	63
<i>Verena Henrich, Erhard Hinrichs and Reinhild Barkey</i>	
Building a standardized Wordnet in the ISO LMF for aeb language	71
<i>Nadia B.M Karmani, Hsan Soussou and Adel M. Alimi</i>	
Java Libraries for Accessing the Princeton Wordnet: Comparison and Evaluation	78
<i>Mark Finlayson</i>	
Concept Space Synset Manager Tool	86
<i>Apurva Nagvenkar, Neha Prabhugaonkar, Venkatesh Prabhu, Ramdas Karmali and Jyoti Pawar</i>	
Use of Sense Marking for Improving WordNet Coverage	95
<i>Neha Prabhugaonkar and Jyoti Pawar</i>	
Building a WordNet for Sinhala	100
<i>Indeewari Wijesiri, Malaka Gallage, Buddhika Gunathilaka, Madhuranga Lakjeewa, Daya Wimalasuriya, Gihan Dias, Rohini Paranavithana and Nisansa de Silva</i>	
Coping with Derivation in the Bulgarian Wordnet	109
<i>Tsvetana Dimitrova, Ekaterina Tarpomanova and Borislav Rizov</i>	
Non-Lexicalized Concepts in Wordnets: A Case Study of English and Hungarian	118
<i>Veronika Vincze and Attila Almási</i>	
Enriching SerbianWordNet and Electronic Dictionaries with Terms from the Culinary Domain	127
<i>Stasa Vujicic Stankovic, Cvetana Krstev and Dusko Vitas</i>	

What implementation and translation teach us: the case of semantic similarity measures in wordnets	133
<i>Marten Postma and Piek Vossen</i>	
Hydra: A Software System for Wordnet	142
<i>Borislav Rizov</i>	
Taking stock of the African Wordnet project: 5 years of development	148
<i>Marissa Griesel and Sonja Bosch</i>	
RuThes Linguistic Ontology vs. Russian Wordnets	154
<i>Natalia Loukachevitch and Boris Dobrov</i>	
One Lexicon, Two Structures: So What Gives?	163
<i>Nabil Gader, Sandrine Ollinger and Alain Polguère</i>	
Automatic Construction of Amharic Semantic Networks from Unstructured Text Using Amharic WordNet	172
<i>Alelgn Tefera and Yaregal Assabie</i>	
Graph Based Algorithm for Automatic Domain Segmentation of WordNet	178
<i>Brijesh Bhatt, Subhash Kunnath and Pushpak Bhattacharyya</i>	
Parse Ranking with Semantic Dependencies and WordNet	186
<i>Xiaocheng Yin, Jung-Jae Kim, Zinaida Pozen and Francis Bond</i>	
Do not do processing, when you can look up: Towards a Discrimination Net for WSD	194
<i>Diptesh Kanojia, Pushpak Bhattacharyya, Raj Dabre, Siddhartha Gunti and Manish Shrivastava</i>	
Elephant Beer and Shinto Gates: Managing Similar Concepts in a Multilingual Database .	201
<i>Martin Benjamin</i>	
Creation of Lexical Relations for IndoWordNet	206
<i>Parteek Kumar, R.K. Sharma and Ashish Narang</i>	
Swesaurus; or, The Frankenstein Approach to Wordnet Construction	215
<i>Lars Borin and Markus Forsberg</i>	
Facilitating Multi-Lingual Sense Annotation: Human Mediated Lemmatizer	224
<i>Dr. Pushpak Bhattacharyya, Ankit Bahuguna, Lavita Talukdar and Bornali Phukan</i>	
VerbNet Workbench.....	232
<i>Indrek Jentson</i>	
A Survey of WordNet Annotated Corpora.....	236
<i>Tommaso Petrolito and Francis Bond</i>	
A Quantitative Analysis of Synset of Assamese WordNet: Its Position and Timeline	246
<i>Shikhar Sarma, Dibyajyoti Sarmah, Ratul Deka, Anup Barman, Jumi Sarmah, Himadri Bharali, Mayashree Mahanta and Umesh Deka</i>	

An Analytical Study of Synonymy in Assamese Language Using WorldNet: Classification and Structure	250
<i>Himadri Bharali, Mayashree Mahanta, Shikhar Kr. Sarma, Utpal Saikia and Dibyajyoti Sarmah</i>	
Assamese WordNet based Quality Enhancement of Bilingual Machine Translation System	256
<i>Anup Barman, Jumi Sarmah and Shikhar Sarma</i>	
Morphosemantic relations between verbs in Croatian WordNet	262
<i>Krešimir Sojat and Matea Srebacic</i>	
News about the Romanian Wordnet	268
<i>Verginica Barbu Mititelu, Stefan Daniel Dumitrescu and Dan Tufiş</i>	
On shape classifiers, their metaphorical extension(s) and wordnet potentials	276
<i>Francesca Quattri</i>	
Leveraging Morpho-semantics for the Discovery of Relations in Chinese Wordnet	283
<i>Shu-Kai Hsieh and Yu-Yun Chang</i>	
Aligning an Italian WordNet with a Lexicographic Dictionary: Coping with limited data..	290
<i>Tommaso Caselli, Carlo Strapparava, Vieu Laure and Guido Vetere</i>	
Terminology in WordNet and in plWordNet	299
<i>Marta Dobrowolska and Stan Szpakowicz</i>	
plWordNet as the Cornerstone of a Toolkit of Lexico-semantic Resources	304
<i>Marek Maziarz, Maciej Piasecki, Ewa Rudnicka and Stan Szpakowicz</i>	
Some structural tests for WordNet with results	313
<i>Ahti Lohk, Heili Orav and Leo Vohandu</i>	
Fusion of Multiple Semantic Networks and Human Association	318
<i>Hitoshi Isahara, Kyoko Kanzaki, Eiko Yamamoto, Takayuki Kuribayashi and Michinaga Otsuka</i>	
Semi-Automatic Extension of Sanskrit Wordnet using Bilingual Dictionary	324
<i>Sudha Bhingardive, Tanuja Ajotikar, Irawati Kulkarni, Malhar Kulkarni and Pushpak Bhattacharyya</i>	
Registers in the System of Semantic Relations in plWordNet	330
<i>Marek Maziarz, Maciej Piasecki, Ewa Rudnicka and Stan Szpakowicz</i>	
IndoWordnet Visualizer: A Graphical User Interface for Browsing and Exploring Wordnets of Indian Languages	338
<i>Devendra Singh Chaplot, Sudha Bhingardive and Pushpak Bhattacharyya</i>	
Towards Building Lexical Ontology via Cross-Language Matching	346
<i>Mamoun Abu Helou, Matteo Palmonari, Mustafa Jarrar and Christiane Fellbaum</i>	
Morphosyntactic discrepancies in representing the adjective equivalent in African WordNet with reference to Northern Sotho	355
<i>Mampaka Lydia Mojapelo</i>	
First steps towards a Predicate Matrix	363
<i>Egoitz Laparra, Maddalen Lopez de Lacalle and German Rigau</i>	

Reducing False Positives in the Construction of Adjective Scales.....	372
<i>Alice Zhang</i>	
Embedding NomLex-BR nominalizations into OpenWordnet-PT.....	378
<i>Alexandre Rademaker, Valeria De Paiva, Gerard de Melo and Livy Maria Real Coelho</i>	
OpenWordNet-PT: A Project Report.....	383
<i>Alexandre Rademaker, Valeria De Paiva, Gerard de Melo, Livy Real and Maira Gatti</i>	
Issues in building English-Chinese parallel corpora with WordNets.....	391
<i>Francis Bond and Shan Wang</i>	
"PolNet - Polish WordNet project: PolNet" 2.0 - a short description of the release	400
<i>Zygmunt Vetulani and Bartłomiej Kochanowski</i>	