

GWC 2023

Proceedings of the 12th Global Wordnet Conference

German Rigau, Francis Bond, Alexandre Rademaker (Eds.)

23–27 Jan, 2023

University of the Basque Country in Donostia-San Sebastian
Basque Country



**Global
WordNet
Association**

**gwc
2023
Donostia**

©2023 Global WordNet Association

ISBN 978-84-09-53956-7

Foreword

The 12th Global Wordnet Conference was celebrated in Donostia-San Sebastián two years after the outbreak of the coronavirus pandemic forced the preceding event to take place online. Fortunately, this year the receding threat from the virus and the slow return to pre-pandemic life allowed scholars and researchers from Africa, America, Asia, and Europe to make the journey to the Basque Country. Reunited in person once more, the conference was an opportunity to greet longtime colleagues and forge new acquaintances. Its varied sessions were interspersed with outings to local landmarks and historical spaces, including a visit to the traditional cider house on an unusually frigid and windswept evening.

These activities were organized by this year's host, HiTZ, the Basque Center for Language and Technology. Many of the center's members have long been active participants in the development of WordNet and its related resources. This includes the steady creation of a Basque WordNet, a notable example of the multilingual nature of wordnets globally.

We received fifty-two submissions, two more than the previous edition. The increase points to the continuing strength of WordNet and its critical importance to Natural Language Processing within the current wave of Large Language Models. Among the forty-three submissions accepted and presented during the conference were studies that discussed the use of Wordnet for improving deep learning, methods to connect wordnets to other ontologies and resources, Wordnet extensions, and wordnets for Latvian, Guarani, Cantonese, and Japanese. Our invited speakers provided histories of NLP and WordNet that took us from their origins to the present day and into the future. Christiane Fellbaum, co-founder and co-president of the Global WordNet Association, presented a retrospective overview of WordNet, while José Camacho, Senior Lecturer at Cardiff University's School of Computer Science and Informatics, discussed the open challenges that exist in word embeddings and language models.

We hope the work collected in this volume will not only encourage further research into wordnets and their place within NLP today, but also serve as a bridge to future advances in the field.

Begoña Altuna, Itziar Aldabe, Xabier Arregi, Itziar Gonzalez-Dios, Aritz Farwell, Esther Miranda.

January 2023

Conference Chairs

- German Rigau, HiTZ Center, University of the Basque Country UPV/EHU
- Francis Bond, Palacký University

Local Organising Committee

- Begoña Altuna, HiTZ Center, University of the Basque Country UPV/EHU
- Itziar Aldabe, HiTZ Center, University of the Basque Country UPV/EHU
- Xabier Arregi, HiTZ Center, University of the Basque Country UPV/EHU
- Itziar Gonzalez-Dios, HiTZ Center, University of the Basque Country UPV/EHU
- Aritz Farwell, HiTZ Center, University of the Basque Country UPV/EHU
- Esther Miranda, HiTZ Center, University of the Basque Country UPV/EHU

Program Committee

- Adam Pease, Articulate Software
- Ales Horak, Masaryk University
- Alexandre Rademaker, IBM Research Brazil and EMAP/FGV
- Bolette Pedersen, University of Copenhagen
- Christiane Fellbaum, Princeton University
- Darja Fiser, University of Ljubljana
- David Lindemann, IWiSt, University of Hildesheim
- Diptesh Kanojia, IIT Bombay
- Eneko Agirre, HiTZ Center, University of the Basque Country UPV/EHU
- Ewa Rudnicka, Wrocław University of Technology
- Francis Bond, Palacký University
- Gerard De Melo, Rutgers University
- German Rigau, HiTZ Center, University of the Basque Country UPV/EHU
- Haldur Oim, University of Tartu

- Heili Orav, University of Tartu
- Hugo Gonçalo-Oliveira, Department of Informatics Engineering of the University of Coimbra
- Janos Csirik, University of Szeged
- John Mccrae, National University of Ireland, Galway
- Kadri Vider, University of Tartu
- Kevin Scannell, Saint Louis University
- Kyoko Kanzaki, Otemon Gakuin University
- Maciej Piasecki, Department of Computational Intelligence, Wroclaw University
- Marten Postma, Vrije Universiteit Amsterdam
- Paul Buitelaar, National University of Ireland, Galway
- Piek Vossen, VU University Amsterdam.
- Sanni Nimb, The Danish Society for Language and Literature
- Shan Wang, The Education University of Hong Kong
- Shu-Kai Hsieh, National Taiwan Normal University
- Sonja Bosch, Department of African Languages, University of South Africa
- Thierry Declerck, DFKI, Saarbruecken
- Tim Baldwin, The University of Melbourne
- Tomaž Erjavec, Dept. of Knowledge Technologies, Jožef Stefan Institute
- Umamaheswari Vasanthakumar, Nanyang Technological University
- Valeria de Paiva, Natural Language and AI Research Laboratory of Nuance Communications, Inc.
- Verginica Mititelu, Romanian Academy Research Institute for Artificial Intelligence

Invited Speakers

- Christiane Fellbaum
- José Camacho Collados

Invited talk

Christiane Fellbaum: 35 years of WordNet: Taking stock and looking ahead

We provide a brief behind-the-scenes look at the early days of WordNet, highlighting its initial motivation and its evolution. Examples of WordNet's current and potential future contributions to research in linguistics and psycholinguistics as well as to a wide range of applications are discussed.

Bio Christiane Fellbaum is a Lecturer with Rank of Professor in the Program in Linguistics and the Computer Science Department at Princeton University. She was educated in Germany, France, and the U.S. and received her Ph.D. in Linguistics from Princeton University. She is a co-developer, with George A. Miller, of the lexical database WordNet, and has been active in WordNet-related research and diverse applications. She has published widely on topics in lexical semantics and computational linguistics.

José Camacho-Collados: Contextualized Embeddings, Word Sense Disambiguation and Open Challenges

Embeddings have been one of the most important topics of interest in Natural Language Processing (NLP) for the past decade. Representing knowledge through low-dimensional vectors that are flexible and easily integrable in modern machine learning models has played a central role in the development of the field. Embedding techniques initially focused on words but the attention soon started to shift to other forms. Recently, contextualized embeddings such as those provided by BERT and similar approaches have taken NLP by storm, providing improvements in many downstream tasks. Unlike static word embeddings, contextualized models can dynamically capture the meaning of a word in context. In this talk, I will explain to what extent this is true, showing the main advantages and limitations of current approaches. I will take word sense disambiguation as a proxy to answer these questions, presenting an overview of the field from a language modelling perspective and discussing open challenges.

Bio Jose Camacho-Collados is a Senior Lecturer and UKRI Future Leaders Fellow at Cardiff University, leading the Cardiff NLP group. Before joining Cardiff University, he completed his PhD in Sapienza University of Rome and was a Google AI PhD Fellow. He has worked on different areas of Natural Language Processing (NLP) with a particular focus on semantics. He is the co-author of the "Embeddings in Natural Language Processing" book and is the current Program Chair of *SEM. In addition to semantics, he is interested in lexical resources and multilinguality, and in the last few years he has worked on developing NLP models specialised in social media, such as those included in the recently released TweetNLP platform.

Table of Contents

Probing Taxonomic and Thematic Embeddings for Taxonomic Information	1
<i>Filip Klubička and John Kelleher</i>	
A WordNet View on Crosslingual Transformers	14
<i>Wondimagegnhue Tufa, Lisa Beinborn and Piek Vossen</i>	
What to Make of make? Sense Distinctions for Light Verbs	25
<i>Julie Kallini and Christiane Fellbaum</i>	
Towards Effective Correction Methods Using WordNet Meronymy Relations	31
<i>Javier Álvez, Itziar Gonzalez-Dios and German Rigau</i>	
On the Acquisition of WordNet Relations in Portuguese from Pretrained Masked Language Models	41
<i>Hugo Gonçalo Oliveira</i>	
Wordnet for Definition Augmentation with Encoder-Decoder Architecture	50
<i>Konrad Wojtasik, Arkadiusz Janz and Maciej Piasecki</i>	
Data Augmentation Method for Boosting Multilingual Word Sense Disambiguation	60
<i>Arkadiusz Janz and Marek Maziarz</i>	
Mapping Wordnets on the Fly with Permanent Sense Keys	67
<i>Eric Kafe</i>	
Linking the Sanskrit WordNet to the Vedic Dependency Treebank: a pilot study	77
<i>Erica Biagetti, Chiara Zanchi and Silvia Luraghi</i>	
StarNet: A WordNet Editor Interface	84
<i>Oğuzhan Kuyrukçu, Ezgi Sanıyar and Olcay Taner Yildiz</i>	
Identifying FrameNet Lexical Semantic Structures for Knowledge Graph Extraction from Financial Customer Interactions	91
<i>Cécile Robin, Atharva Kulkarni and Paul Buitelaar</i>	
Some Considerations in the Construction of a Historical Language WordNet	101
<i>Fahad Khan, John P. McCrae, Francisco Javier Minaya Gómez, Rafael Cruz González and Javier E. Díaz-Vera</i>	
Hidden in Plain Sight: Can German Wiktionary and Wordnets Facilitate the Detection of Antithesis?	106
<i>Ramona Kuehn, Jelena Mitrović and Michael Granitzer</i>	
How do We Treat Systematic Polysemy in Wordnets and Similar Resources? – Using Human Intuition and Contextualized Embeddings as Guidance	117
<i>Nathalie Sørensen, Sanni Nimb and Bolette Pedersen</i>	
The Romanian Wordnet in Linked Open Data Format	127
<i>Elena Irimia and Verginica Mititelu</i>	
Combining WordNets with Treebanks to study idiomatic language: A pilot study on Rigvedic formulas through the lenses of the Sanskrit WordNet and the Vedic Treebank	133
<i>Luca Brigada Villa, Erica Biagetti, Riccardo Ginevra and Chiara Zanchi</i>	
Word Sense Disambiguation Based on Iterative Activation Spreading with Contextual Embeddings for Sense Matching	140

Arkadiusz Janz and Maciej Piasecki

Documenting the Open Multilingual Wordnet	150
<i>Francis Bond, Michael Wayne Goodman, Ewa Rudnicka, Luis Morgado da Costa, Alexandre Rademaker and John P. McCrae</i>	
Mapping GermaNet for the Semantic Web using OntoLex-Lemon	158
<i>Claus Zinn, Marie Hinrichs and Erhard Hinrichs</i>	
Incorporating prepositions in the BulTreeBank WordNet	167
<i>Zara Kancheva</i>	
Are there just WordNets or also SignNets?	172
<i>Ineke Schuurman, Thierry Declerck, Caro Brosens, Margot Janssens, Vincent Vandeghinste and Bram Vanroy</i>	
The Japanese Wordnet 2.0	179
<i>Francis Bond and Takayuki Kuribayashi</i>	
Latvian WordNet	187
<i>Peteris Paikens, Agute Klints, Ilze Lokmane, Lauma Pretkalniņa, Laura Rituma, Madara Stāde and Laine Strankale</i>	
Initial Experiments for Building a Guarani WordNet	197
<i>Luis Chiruzzo, Marvin Agüero-Torales, Aldo Alvarez and Yliana Rodríguez</i>	
A CCGbank for Turkish: From Dependency to CCG	205
<i>Aslı Kuzgun, Oğuz Kerem Yıldız and Olcay Taner Yildiz</i>	
Reusing the Danish WordNet for a New Central Word Register for Danish - a Project Report	214
<i>Bolette Pedersen, Sanni Nimb, Nathalie Sørensen, Sussi Olsen, Ida Flörke and Thomas Troelsgård</i>	
Recent Developments in BTB-WordNet	220
<i>Kiril Simov and Petya Osenova</i>	
Lexicalised and non-lexicalized multi-word expressions in WordNet: a cross-encoder approach	228
<i>Marek Maziarz, Lukasz Grabowski, Tadeusz Piotrowski, Ewa Rudnicka and Maciej Piasecki</i>	
Towards an RDF Representation of the Infrastructure consisting in using Wordnets as a conceptual Interlingua between multilingual Sign Language Datasets	235
<i>Thierry Declerck, Thomas Troelsgård and Sussi Olsen</i>	
Semantic Parsing and Sense Tagging the Princeton WordNet Gloss Corpus	243
<i>Alexandre Rademaker, Abhishek Basu and Rajkiran Veluri</i>	
Context-Gloss Augmentation for Improving Arabic Target Sense Verification	254
<i>Sanad Malaysha, Mustafa Jarrar and Mohammed Khalilia</i>	
The Open Cantonese Sense-Tagged Corpus	263
<i>Joanna Sio and Luis Morgado Da Costa</i>	
Correcting Sense Annotations Using Wordnets and Translations	269
<i>Arnob Mallik and Grzegorz Kondrak</i>	
A Benchmark and Scoring Algorithm for Enriching Arabic Synonyms	274
<i>Sana Ghanem, Mustafa Jarrar, Radi Jarrar and Ibrahim Bounhas</i>	
Expanding the Conceptual Description of Verbs in WordNet with Semantic and Syntactic Information	284

<i>Ivelina Stoyanova and Svetlozara Leseva</i>	
An Experiment: Finding Parents for Parentless Synsets by Means of CILI	295
<i>Ahti Lohk, Martin Rebane and Heili Orav</i>	
Extending the usage of adjectives in the Zulu AfWN	303
<i>Laurette Marais and Laurette Pretorius</i>	
Linking SIL Semantic Domains to Wordnet and Expanding the Abui Wordnet through Rapid Word Collection Methodology	315
<i>Luis Morgado Da Costa, František Kratochvíl, George Saad, Benidiktus Delpada, Daniel Simon Lanma, Francis Bond, Natálie Wolfová and A.L. Blake</i>	
Wordnet-oriented recognition of derivational relations	325
<i>Wiktor Walentynowicz and Maciej Piasecki</i>	
What do Language Models know about word senses? Zero-Shot WSD with Language Models and Domain Inventories	331
<i>Oscar Sainz, Oier Lopez de Lacalle, Eneko Agirre and German Rigau</i>	
Resolving Multiple Hyperonymy	343
<i>Svetla Koeva and Dimitar Hristov</i>	
Towards the integration of WordNet into ClinIDMap	352
<i>Elena Zotova, Montse Cuadros and German Rigau</i>	
Connecting Multilingual Wordnets: Strategies for Improving ILI Classification in OdeNet	363
<i>Melanie Siegel and Johann Bergh</i>	