GWC 2021

Proceedings of the 11th Global Wordnet Conference

Sonja Bosch, Christiane Fellbaum, Marissa Griesel,
Alexandre Rademaker and Piek Vossen (Eds.)

18–21 Jan, 2021
South African Centre for Digital Language Resources (SADiLaR)
Potchefstroom, South Africa

©2021 Global WordNet Association

ISBN 978-9-464027-31-0

Foreword

We are excited to hold the 11^{th} Global Wordnet conference on the continent where our human ancestors first created language tens of millions of years ago. South Africa today is home to eleven official and at least twenty-five other languages, and some joined the community of wordnet builders more than a decade ago when African Wordnet was launched. While the global pandemic has prevented us from meeting in person and forcing us to forego coffee and sightseeing breaks, the virtual format allows everyone to participate without incurring travel costs, and jetlag when attending talks outside of one's time zone is merely optional.

We received fifty submissions, and forty-one papers will be presented by colleagues from all continents except Antarctica. We will hear about wordnets covering languages that are new to our community (Uzbek, ancient Indo-European languages, taboo language), new approaches to the automatic construction of wordnets, enhancements of the "classic" WordNet model with additional relations and semantic information, crosslingual wordnet alignment, tools and applications for NLP tasks. Our invited speaker from Palestine, a short transcontinental hop across the Sinai, highlights the important distinctions between a Wordnet and an ontology, showing how ontology engineering can inform wordnet construction.

We are grateful to the South African Centre for Digital Language Resources (SADiLaR), without whose sponsorship and hosting this conference could not have taken place. Thanks go to the local organizers, who volunteered their time and effort, and to the members of the Program Committee, who read and reviewed submissions.

Christiane Fellbaum

Piek Vossen

Sonja Bosch

Marissa Griesel

Mampaka Lydia Mojapelo

Juan Steyn

Elsabé Taljard

Liané van den Bergh

January 2021

Conference Chairs

- Christiane Fellbaum (Princeton University)
- Piek Vossen (Vrije Universiteit Amsterdam)

Local Organising Committee

- Sonja Bosch – University of South Africa (UNISA)
- Mampaka Lydia Mojapelo – University of South Africa (UNISA)
- Marissa Griesel – University of South Africa (UNISA)
- Elsabé Taljard – University of Pretoria (UP)
- Juan Steyn – SADiLaR
- Liané van den Bergh – SADiLaR

Program Committee:

- Eneko Agirre (University of the Basque Country)
- Sina Ahmadi (Insight Centre for Data Analytics)
- Timothy Baldwin (The University of Melbourne)
- Francis Bond (Nanyang Technological University)
- Sonja Bosch (Department of African Languages, University of South Africa)
- Paul Buitelaar (Insight Centre for Data Analytics, National University of Ireland Galway)
- Bharathi Raja Chakravarthi (Insight Centre for Data Analytics, National University of Ireland, Galway)
- Janos Csirik (University of Szeged)
- Gerard de Melo (Rutgers University)
- Valeria de Paiva (Samsung Research America and University of Birmingham)
- Thierry Declerck (DFKI GmbH)
- Bento C. Dias-Da-Silva (UNESP)
- Umamaheswari E (Research Fellow NTU Singapore)

- Christiane Fellbaum (Princeton University)
- Leonel Figueiredo de Alencar (UNIVERSIDADE FEDERAL DO CEARÁ)
- Hugo Gonçalo Oliveira (University of Coimbra)
- Ales Horak (Masaryk University, Faculty of Informatics)
- Shu-Kai Hsieh (National Taiwan Normal University)
- Filip Ilievski (Information Sciences Institute, University of Southern California)
- Diptesh Kanojia (IIT Bombay)
- Kyoko Kanzaki Toyohashi (University of Technology)
- Shikhar Kr. Sarma (Gauhati University)
- David Lindemann (UPV-EHU University of the Basque Country)
- Ahti Lohk (Tallinn University of Technology)
- John P. McCrae (National University of Ireland, Galway)
- Verginica Mititelu (Romanian Academy Research Institute for Artificial Intelligence)
- Luis Morgado Da Costa (Nanyang Technological University)
- Sanni Nimb (Det Danske Sprog-og Litteraturselskab, DSL)
- Sussi Olsen (University of Copenhagen, Centre for Language Technology)
- Heili Orav (University of Tartu)
- Bolette Pedersen (University of Copenhagen)
- Maciej Piasecki (Department of Computational Intelligence, Wroclaw University of Science and Technology)
- Marten Postma (Vrije Universiteit Amsterdam)
- Alexandre Rademaker (IBM Research Brazil and EMAp/FGV)
- German Rigau (IXA Group, UPV/EHU)
- Ewa Rudnicka (Wrocław University of Technology)
- Kevin Scannell (Saint Louis University)
- Pia Sommerauer (Vrije Universiteit Amsterdam)
- Kadri Vider (University of Tartu)
- Piek Vossen (Vrije Universiteit Amsterdam)
- Shan Wang (University of Macau)

Invited Speaker

Mustafa Jarrar, Birzeit University, Palestine

Invited talk

Mustafa Jarrar: Linguistic Ontologies and Wordnets

Wordnets play an important role in understanding and retrieving unstructured information, especially in NLP and IR tasks. Their importance is also increasing to support managing and retrieving of structured data in new areas, such as Knowledge Graphs, multilingual Big Data, and medical informatics. Such new needs are demanding wordnets to be formal and play the role of ontologies.

The difference between wordnets and ontologies might not be obvious, especially because both have similar structures, e.g. considering synsets as concepts and hyponyms as subsumptions. However, synsets in wordnets are linguistically motivated concepts (i.e. units of thoughts), while concepts in ontologies are classes of instances. Additionally, subsumption is a subset relation, in the extensional or intensional sense, rather than a linguistic general-specific relationship. Furthermore, ontologies are typically application-specific rich axiomatizations, while wordnets are general-purpose mental lexicons, thus axiomatizing them would be a rigidification.

This talk will discuss the notion of linguistic ontology, which can play the role of being a wordnet and an ontology at the same time. The talk will also discuss what can be learned from the ontology engineering literature to build wordnets with ontologically and formally cleaner content.

The second part of the talk will present the Arabic Ontology, which is an Arabic wordnet built with formal and ontological analysis in mind. The ontology is represented in a similar structure as wordnets, and is fully mapped to the Princeton Wordnet, as well as with the WikiData knowledge graph and with many Arabic-multilingual lexicons. The ontology is being built at Birzeit University, in Palestine, and it is available at https://ontology.birzeit.edu/concept/293198.

Table of Contents

On Universal Colexifications 1
Hongchang Bao, Bradley Hauer and Grzegorz Kondrak

UZWORDNET: A Lexical-Semantic Database for the Uzbek Language 8
Alessandro Agostini, Timur Usmanov, Ulugbek Khamdamov, Nilufar Abdurakhmonova and Mukhammadsaid Mamasaidov

Practical Approach on Implementation of WordNets for South African Languages 20
Tshephisho Joseph Sefara, Tumisho Billson Mokgonyane and Vukosi Marivate

Homonymy and Polysemy Detection with Multilingual Information 26
Amir Ahmad Habibi, Bradley Hauer and Grzegorz Kondrak

Taboo Wordnet 36
Francis Bond and Merrick Yeu Herng Choo

Ask2Transformers: Zero-Shot Domain labelling with Pretrained Language Models 44
Oscar Sainz and German Rigau

Discriminating Homonymy from Polysemy in Wordnets: English, Spanish and Polish Nouns . . . 53
Arkadiusz Janz and Marek Maziarz

Implementing ASLNet V1.0: Progress and Plans 63
Colin Lualdi, Elaine Wright, Jack Hudson, Naomi Caselli and Christiane Fellbaum

Monolingual Word Sense Alignment as a Classification Problem 73
Sina Ahmadi and John P. McCrae

Extraction of Common-Sense Relations from Procedural Task Instructions using BERT 81
Viktor Losing, Lydia Fischer and Jörg Deigmöller

The GlobalWordNet Formats: Updates for 2020 91
John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka and Luis Morgado Da Costa

Intrinsically Interlingual: The Wn Python Library for Wordnets 100
Michael Wayne Goodman and Francis Bond

Semantic Analysis of Verb-Noun Derivation in Princeton WordNet 108
Verginica Mititelu, Svetlozara Leseva and Ivelina Stoyanova

Building the Turkish FrameNet 118
Büşra Marşan, Neslihan Kara, Merve Özçelik, Bilge Nas Arıcan, Neslihan Cesur, Aslı Kuzgun, Ezgi Sanıyar, Oğuzhan Kuyrukçu and Olcay Taner Yildiz

Evaluation of Taxonomy Enrichment on Diachronic WordNet Versions 126
Irina Nikishina, Natalia Loukachevitch, Varvara Logacheva and Alexander Panchenko

A (Non)-Perfect Match: Mapping plWordNet onto PrincetonWordNet 137
Ewa Rudnicka, Wojciech Witkowski and Maciej Piasecki

Persian SemCor: A Bag of Word Sense Annotated Corpus for the Persian Language 147
Hossein Rouhizadeh, Mehrnoush Shamsfard, Mahdi Dehghan and Masoud Rouhizadeh

HisNet: A Polarity Lexicon based on WordNet for Emotion Analysis 157

Merve Özçelik, Bilge Nas Arıcan, Özge Bakay, Elif Sarmış, Özlem Ergelen, Nilgün Güler Bayezit and Olcay Taner Yıldız

Turkish WordNet KeNet 166
Özge Bakay, Özlem Ergelen, Elif Sarmış, Selin Yıldırım, Bilge Nas Arıcan, Atilla Kocabalcıoğlu, Merve Özçelik, Ezgi Sanıyar, Oğuzhan Kuyrukçu, Begüm Avar and Olcay Taner Yıldız

Enriching plWordNet with morphology 175
Agnieszka Dziob and Wiktor Walentynowicz

Towards Expanding WordNet with Conceptual Frames 182
Koeva Svetla

OdeNet: Compiling a GermanWordNet from other Resources 192
Melanie Siegel and Francis Bond

Comparing Similarity of Words Based on Psychosemantic Experiment and RuWordNet 199
Valery Solovyev and Natalia Loukachevitch

Text Document Clustering: Wordnet vs. TF-IDF vs. Word Embeddings 207
MichałMarcińczuk, Mateusz Gniewkowski, Tomasz Walkowiak and Marcin Będkowski

Extracting Synonyms from Bilingual Dictionaries 215
Mustafa Jarrar, Eman Naser, Muhammad Khalifa and Khaled Shaalan

Neural Language Models vs Wordnet-based Semantically Enriched Representation in CST Relation Recognition 223
Arkadiusz Janz, Maciej Piasecki and Piotr Wątorski

What is on Social Media that is not in WordNet? A Preliminary Analysis on the TwitterAAE Corpus 234
Cecilia Domingo, Tatiana Gonzalez-Ferrero and Itziar Gonzalez-Dios

Creating Domain Dependent Turkish WordNet and SentiNet 243
Bilge Nas Arıcan, Merve Özçelik, Deniz Baran Aslan, Elif Sarmış, Selen Parlar and Olcay Taner Yıldız

Towards a Linking between WordNet and Wikidata 252
John P. McCrae and David Cillessen

Toward the creation of WordNets for ancient Indo-European languages 258
Erica Biagetti, Chiara Zanchi and William Michael Short

DanNet2: Extending the coverage of adjectives in DanNet based on thesaurus data (project presentation) 267
Sanni Nimb, Bolette Pedersen and Sussi Olsen

Teaching Through Tagging —Interactive Lexical Semantics 273
Francis Bond, Andrew Devadason, Melissa Rui Lin Teo and Luís Morgado da Costa

Towards the Addition of Pronunciation Information to Lexical Semantic Resources 284
Thierry Declerck and Lenka Bajčetić

Testing agreement between lexicographers: A case of homonymy and polysemy 292
Marek Maziarz, Francis Bond and Ewa Rudnicka