The Workshop on Cognitive Aspects of the Lexicon (CogALex@LREC-COLING 2024)

Workshop Proceedings

Editors
Michael Zock, Emmanuele Chersoni, Yu-Yin Hsu and
Simon De Deyne

20 May, 2024
Torino, Italia

Proceedings of the Workshop on Cognitive Aspects of the Lexicon (CogaLex@LREC-COLING 2024)

Copyright ELRA Language Resources Association (ELRA), 2024
These proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-45-6
ISSN 2951-2093 (COLING); 2522-2686 (LREC)

Jointly organized by the ELRA Language Resources Association
and the International Committee on Computational Linguistics

Message from the Program Chairs

Words are the exchange money for revealing our thoughts. While a single word may capture a world of knowledge in a nutshell, its combination with other words may allow for even more, support communication, memorization, and even thinking, an activity that consists largely of perceiving links, and combining concepts for which we have created words

Words can entertain, reveal, or hide. Sharper than razor blades, they can also make us laugh or cry. No doubt, words are important. They are everywhere and multifarious, and different disciplines have contributed to explain how they are learned, represented, organized, or used. While we must continue to draw on expertise from various fields (linguistics, psychology, etc.), it is essential to facilitate the communication of the most recent developments in each of them. For example, widely used techniques like association thesauri, complex graphs, or large language models (LLMs) have not been created by linguists but by psychologists or specialists in AI (mathematicians, statisticians, and computer scientists). The knowledge in each of these areas is growing at an exceptional rate, and what used to be true yesterday may not be so anymore today or tomorrow. This workshop aims to take stock of these developments and support the continuing exchange between researchers across disciplines. These are but some of the motivations for organizing this kind of workshop. For more details concerning the landscape, the diversity of problems and solutions, see our homepage: `https://sites.google.com/view/cogalex-viii-2024/home`.

This is the 8th CogALex workshop. Following the precursor 'Enhancing and Using Electronic Dictionaries' (2004, COLING, Geneva, Switzerland), there have been seven similar events, though with a stronger focus on the 'Cognitive Aspects of the Lexicon', hence the name and acronym of this workshop, CogALex:

- CogALex-I (2008, COLING, Manchester, UK)
- CogALex-II (2010, COLING, Beijing, China)
- CogALex-III (2012, COLING, Mumbai, India)
- CogALex-IV (2014, COLING, Dublin, Ireland)
- CogALex-V (2016, COLING, Osaka, Japan)
- CogALex-VI (2020, COLING, Barcelona, Spain)
- CogALex-VII (2022, AACL-IJCNLP, Taipei, Taiwan)

As every organizer knows, no workshop can be held without the help of competent and devoted reviewers (`https://sites.google.com/view/cogalex-viii-2024/home/programme-committee`). We would like to take this opportunity to thank all of them for their hard work. Last, but not least we would like to thank Gilles-Maurice de Schryver who was so kind to accept our invitation as an invited speaker. Being an experienced professional, we are particularly keen to hear his thoughts about the potential uses of LLMs for dictionary making (`https://sites.google.com/view/cogalex-viii-2024/home/invited-speaker`).

In this edition of the workshop, we have received 24 submissions and accepted 19 of them for presentation (79% of acceptance rate).

One last point to the authors. We have received an unusually high number of excellent papers, which would easily fit a full day workshop. Unfortunately only half a day workshop could be scheduled this year, which means that we could not have enough oral presentation slots available reflecting the high quality of submissions. We have tried to balance the time for oral and poster presentations to do justice to both. We'd like to thank all authors for submitting their best work and look forward to meeting you in Turin.

The Organizers
Michael Zock, Emmanuele Chersoni, Yu-Yin Hsu & Simon De Deyne

Organizing Committee

Michael Zock (CNRS & Aix-Marseille Université, France)
Emmanuele Chersoni (The Hong Kong Polytechnic University, Hong Kong, China)
Simon De Deyne (University of Melbourne, Australia)
Yu-Yin Hsu (The Hong Kong Polytechnic University, Hong Kong, China)

Program Committee

Verginica Barbu Mititelu (RACAI, Romania)
Chris Biemann (University of Hamburg, Germany)
Jose Camacho-Collados (Cardiff University, United Kingdom))
Ya-Ning Chang (University of Cambridge, United Kingdom)
Kenneth Church (Northeastern University, United States)
Rodolfo Delmonte (Ca' Foscari University of Venice, Italy)
Olivier Ferret (Institut CEA List, France)
Thierry Fontenelle (European Investment Bank, Luxembourg)
Thomas François (Université Catholique de Louvain, Belgium)
Markus J. Hofmann (Bergische Universität Wuppertal, Germany)
Shu-Kai Hsieh (National Taiwan University, Taiwan)
Ignacio Iacobacci (Huawei Noah's Ark Lab, United Kingdom)
Mathieu Lafourcade (Université de Montpellier, France)
Phillippe Langlais (Université de Montréal, Canada)
Michael S. Vitevitch (University of Kansas, United States)
Adam Pease (Articulate Software, United States)
Alessandro Raganato (University of Milano-Bicocca, Italy)
Giulia Rambelli (University of Bologna, Italy)
Didier Schwab (Université Grenoble Alpes, France)
Cynthia Siew (National University of Singapore, Singapore)
Massimo Stella (University of Trento, Italy)
Gilles Sérasset (Université Grenoble Alpes, France)
Carole Tiberius (Leiden University, Netherlands)
Yu Wang (The Hong Kong Polytechnic University, Hong Kong, China)
Werner Winiwarter (University of Vienna, Austria)
Pierre Zweigenbaum (Université Paris-Saclay, France)

Table of Contents

CLAVELL - Cognitive Linguistic Annotation and Visualization Environment for Language Learning
Werner Winiwarter 1

Individual Text Corpora Predict Openness, Interests, Knowledge and Level of Education
Markus J. Hofmann, Markus T. Jansen, Christoph Wigbels, Benny Briesemeister and Arthur M. Jacobs 14

An Empirical Study on Vague Deictic Temporal Adverbials
Svenja Kenneweg, Brendan Balcerak Jackson, Joerg Deigmoeller, Julian Eggert and Philipp Cimiano 26

Symbolic Learning of Rules for Semantic Relation Types Identification in French Genitive Postnominal Prepositional Phrases
Hani Guenoune and Mathieu Lafourcade 32

How Human-Like Are Word Associations in Generative Models? An Experiment in Slovene
Špela Vintar, Mojca Brglez and Aleš Žagar 42

Idiom Complexity in Apple-Pie Order: The Disentanglement of Decomposability and Transparency
Irene Pagliai 49

What GPT-4 Knows about Aspectual Coercion: Focused on "Begin the Book"
Seohyun IM and Chungmin Lee 56

Can GPT-4 Recover Latent Semantic Relational Information from Word Associations? A Detailed Analysis of Agreement with Human-annotated Semantic Ontologies.
Simon De Deyne, Chunhua Liu and Lea Frermann 68

What's in a Name? Electrophysiological Differences in Processing Proper Nouns in Mandarin Chinese
Bernard A. J. Jap, Yu-Yin Hsu, Lavinia Salicchi and Yu Xi Li 79

Cross-Linguistic Processing of Non-Compositional Expressions in Slavic Languages
Iuliia Zaitova, Irina Stenger, Muhammad Umer Butt and Tania Avgustinova 86

Using Language Models to Unravel Semantic Development in Children's Use of Perception Verbs
Bram van Dijk, Max J. van Duijn, Li Kloostra, Marco Spruit and Barend Beekhuizen 98

Representing Abstract Concepts with Images: An Investigation with Large Language Models
Ludovica Cerini, Alessandro Bondielli and Alessandro Lenci 107

Big-Five Backstage: A Dramatic Dataset for Characters Personality Traits & Gender Analysis
Vadim A. Porvatov, Carlo Strapparava and Marina Tiuleneva 114

Interaction of Semantics and Morphology in Russian Word Vectors
Yulia Zinova, Ruben van de Vijver and Anastasia Yablokova 120

Listen, Repeat, Decide: Investigating Pronunciation Variation in Spoken Word Recognition among Russian Speakers
Vladislav Ivanovich Zubov and Elena Riekhakaynen 129

The Mental Lexicon of Communicative Fragments and Contours: The Remix N-gram Method
Emese K. Molnár and Andrea Dömötör 133

Three Studies on Predicting Word Concreteness with Embedding Vectors
Michael M. Flor 140

Combining Neo-Structuralist and Cognitive Approaches to Semantics to Build Wordnets for Ancient Languages: Challenges and Perspectives
Erica Biagetti, Martina Giuliani, Silvia Zampetta, Silvia Luraghi and Chiara Zanchi 151

SensoryT5: Infusing Sensorimotor Norms into T5 for Enhanced Fine-grained Emotion Classification
Yuhan Xia, Qingqing Zhao, Yunfei Long, Ge Xu and Jia Wang 162

Workshop Program

Monday, May 20, 2024

9:00–9:15 *Introduction*
Michael Zock

9:15–10:15 Keynote Speech

9:15–10:15 *Fine-tuning LLMs for lexicography*
Gilles-Maurice de Schryver

10:15–10:40 Oral Presentations 1

CLAVELL - Cognitive Linguistic Annotation and Visualization Environment for Language Learning
Werner Winiwarter

10:40–11:00 Coffee Break

11:00–12:00 Poster Session

Individual Text Corpora Predict Openness, Interests, Knowledge and Level of Education
Markus J. Hofmann, Markus T. Jansen, Christoph Wigbels, Benny Briesemeister and Arthur M. Jacobs

An Empirical Study on Vague Deictic Temporal Adverbials
Svenja Kenneweg, Brendan Balcerak Jackson, Joerg Deigmoeller, Julian Eggert and Philipp Cimiano

Symbolic Learning of Rules for Semantic Relation Types Identification in French Genitive Postnominal Prepositional Phrases
Hani Guenoune and Mathieu Lafourcade

How Human-Like Are Word Associations in Generative Models? An Experiment in Slovene
Špela Vintar, Mojca Brglez and Aleš Žagar

Idiom Complexity in Apple-Pie Order: The Disentanglement of Decomposability and Transparency
Irene Pagliai

Monday, May 20, 2024 (continued)

What GPT-4 Knows about Aspectual Coercion: Focused on "Begin the Book"
Seohyun IM and Chungmin Lee

Can GPT-4 Recover Latent Semantic Relational Information from Word Associations? A Detailed Analysis of Agreement with Human-annotated Semantic Ontologies.
Simon De Deyne, Chunhua Liu and Lea Frermann

What's in a Name? Electrophysiological Differences in Processing Proper Nouns in Mandarin Chinese
Bernard A. J. Jap, Yu-Yin Hsu, Lavinia Salicchi and Yu Xi Li

Cross-Linguistic Processing of Non-Compositional Expressions in Slavic Languages
Iuliia Zaitova, Irina Stenger, Muhammad Umer Butt and Tania Avgustinova

Using Language Models to Unravel Semantic Development in Children's Use of Perception Verbs
Bram van Dijk, Max J. van Duijn, Li Kloostra, Marco Spruit and Barend Beekhuizen

Representing Abstract Concepts with Images: An Investigation with Large Language Models
Ludovica Cerini, Alessandro Bondielli and Alessandro Lenci

Big-Five Backstage: A Dramatic Dataset for Characters Personality Traits & Gender Analysis
Vadim A. Porvatov, Carlo Strapparava and Marina Tiuleneva

Interaction of Semantics and Morphology in Russian Word Vectors
Yulia Zinova, Ruben van de Vijver and Anastasia Yablokova

Listen, Repeat, Decide: Investigating Pronunciation Variation in Spoken Word Recognition among Russian Speakers
Vladislav Ivanovich Zubov and Elena Riekhakaynen

The Mental Lexicon of Communicative Fragments and Contours: The Remix N-gram Method
Emese K. Molnár and Andrea Dömötör

Three Studies on Predicting Word Concreteness with Embedding Vectors
Michael M. Flor

Monday, May 20, 2024 (continued)

12:00–12:50 **Oral Presentations 2**

Combining Neo-Structuralist and Cognitive Approaches to Semantics to Build Wordnets for Ancient Languages: Challenges and Perspectives
Erica Biagetti, Martina Giuliani, Silvia Zampetta, Silvia Luraghi and Chiara Zanchi

SensoryT5: Infusing Sensorimotor Norms into T5 for Enhanced Fine-grained Emotion Classification
Yuhan Xia, Qingqing Zhao, Yunfei Long, Ge Xu and Jia Wang

12:50–13:00 *Conclusion*

Michael Zock