CogALex-VII

The Workshop on Cognitive Aspects of the Lexicon

Proceedings of the Workshop

November 20, 2022

Order copies of this and other ACL proceedings from:

Preface

Being ubiquitous and versatile, words are important for many tasks: sharing, storing, processing and accessing information. They are vital not only for communication, the acquisition, structuring (categorization), maintenance and access of knowledge, but also for thinking: problem-solving, link discovery, detection of similarities, metaphors and analogies. This being so, one may wonder how to build tools supporting the learning of words (memorization) and the various types of usage (access/navigation). Alas, the answer is not easy, as it depends on many factors: the task or goal (production/reception), the user's background (age, expertise), the momentary cognitive state (information available at the onset of the search), the material support (book, computer), etc. Obviously, words in books, computers, and the human brain are not the same. Being aware of this, different communities (linguists, lexicographers, psychologists) have focused on different aspects: representation and organization of words in dictionaries, creation of tools supporting navigation or conceptual search (thesauri), time course of word access, etc. The situation is complex, as the respective views, methods and research goals are not quite the same, all the more as they have changed considerably over time.

For example, rather than considering the lexicon as a static entity, where discrete units (words) are organized alphabetically (database view), dictionaries are now viewed dynamically, i.e., as lexical graphs, whose entities are linked in various ways (topical relations; associations) and whose link weights may vary over time. Also, lexicographers view words as products, i.e., holistic entities, while psychologists and neuroscientists view them as processes. Words are decomposed, and their synthesis requires activation of various parts of our brain, each one of them being dedicated to a specific part (meaning, form, sound) of the final form. The normal time course starts with some input (concrete object in the real world, or, more or less clear ideas, abstract concepts) leading then, 'little by little', to an output (word form). All this is achieved in milliseconds, as normal discourse consists in the production of two to three words per second.

Computational linguists have their own ways to look at words, which also have changed quite a bit over time. Discrete count-based vector representations have successively been replaced by continuous vectors (i.e., word embeddings), and then by language-model-based contextualized representations which outperform the static models (including word-embeddings) in a broad range of tasks.

As one can see, different communities look at words from different angles, which can be an asset and a problem. It is an asset, as complementary views may help us to broaden and deepen our understanding of this fundamental cognitive resource. Yet, this diversity of perspectives can also be a problem, in particular in a dynamic field like ours that is so rapidly moving on. Hence, it becomes harder and harder for everyone, including experts, to remain fully informed about the latest changes (state of the art). This is one of the reasons why we organize this workshop. More precisely, our goal is not only to keep people informed without getting them crushed by the information glut, but also to help them to perceive clearly what is new, relevant, and hence important. Last, but not least, we would like to connect people from different communities in the hope that this may help them to gain new insights or inspiration.

This is the 7th edition of CogALex, the first one not to be associated with COLING, as CogALex-VII is part of AACL-IJCNLP 2022. We have received 14 submissions, out of which we have selected 4 for oral presentation, and 6 for posters. This amounts to a submission rate of 71% (28,5% for the papers, 42,5% for the posters). In addition, we have an invited speaker, Massimo Stella from the CogNosco Lab of the University of Exeter (UK). His talk — Multiplex networks and AI unveil the influence of the mental lexicon on picture naming and its failures by people struck with aphasia.— fits our goal perfectly well, as it demonstrates the potential of graph theory to shed some light on the structure and evolution of the mental lexicon. We would like to express our thanks to him for having accepted to be our invited speaker. Also, sincerest thanks to all the members of the Programme Committee. Their expertise was invaluable

to ensure a good selection of papers despite the tight schedule. Their reviews were helpful not only for us to make the decisions, but also for the authors, helping them to strengthen their work. We hope that the work presented here will inspire you, generate fruitful discussions, and possibly lead to new ideas, insights, and collaborations.

The CogALex-VII Workshop Chairs

Organizing Committee

Michael Zock (Aix-Marseille University)
Emmanuele Chersoni (The Hong Kong Polytechnic University)
Yu-Yin Hsu (The Hong Kong Polytechnic University)
Enrico Santus (Bloomberg LP)

Program Committee

Chris Biemann (University of Hamburg)
Marc Brysbaert (Ghent University)
Nichol Castro (University at Buffalo)
Yan Cong (Feinstein Institute for Medical Research)
Simon De Deyne (University of Melbourne)
Rodolfo Delmonte (University Ca' Foscari of Venice)
Luis Espinosa Anke (Cardiff University)
Olivier Ferret (CEA List)
Jinghang Gu (The Hong Kong Polytechnic University)
Eduard Hovy (Carnegie Mellon University)
Shu-Kai Hsieh (National Taiwan University)
Philippe Langlais (University of Montréal)
Alessandro Lenci (University of Pisa)
Vito Pirrelli (ILC-CNR Pisa)
Alain Polguère (University of Lorraine)
Jakob Prange (The Hong Kong Polytechnic University)
Alessandro Raganato (University of Milan-Bicocca)
Lavinia Salicchi (The Hong Kong Polytechnic University)
Sabine Schulte im Walde (University of Stuttgart)
Didier Schwab (Grenoble Alpes University)
Cynthia Siew (National University of Singapore)
Gilles Sérasset (Grenoble Alpes University)

Table of Contents

Conference Program

November 20, 2022

17:00–17:15   **(UTC +8, Taipei time) Introduction**

17:15–18:15   **Keynote Talk**

Multiplex Networks and AI Unveil the Influence of the Mental Lexicon on Picture Naming and its Failures by People Struck with Aphasia
Massimo Stella (University of Exeter)

18:15–18:30   **Short Break**

18:30–19:30   **Oral Presentations 1: Cognitive Modeling and Psycholinguistics**

Patterns of Text Readability in Human and Predicted Eye Movements
Nora Hollenstein, Itziar Gonzalez-Dios, Lisa Beinborn and Lena Jäger

(In)Alienable Possession in Mandarin Relative Clauses
Deran Kong and Yu-Yin Hsu

19:30–20:30   **Long Break**

20:30–21:45   **Poster Session**

Do Age of Acquisition and Orthographic Transparency Have the Same Effects in Different Modalities?
Mohammad Momenian

CAT ManyNames: A New Dataset for Object Naming in Catalan
Mar Domínguez Orfila, Maite Melero Nogués and Gemma Boleda Torrent

Finetuning Latin BERT for Word Sense Disambiguation on the Thesaurus Linguae Latinae
Piroska Lendvai and Claudia Wick

Putting WordNet's Dictionary Examples in the Context of Definition Modelling: An Empirical Analysis
Fatemah Almeman and Luis Espinosa Anke

Exploring Nominal Coercion in Semantic Spaces with Static and Contextualized Word Embeddings
Chenxin LIU and Emmanuele Chersoni

A Frame-Based Model of Inherent Polysemy, Copredication and Argument Coercion
Chen Long, Laura Kallmeyer and Rainer Osswald

21:45–22:45   **Oral Presentations 2: Lexical and Distributional Semantics**

VISCOSE - a Kanji Dictionary Enriched with VISual, COmpositional, and SEmantic Information
Werner Winiwarter and Bartholomäus Wloka

Compositionality as an Analogical Process: Introducing ANNE
Giulia Rambelli, Emmanuele Chersoni, Philippe Blache and Alessandro Lenci

November 20, 2022 (continued)

22:45–23:00   Conclusive Remarks