COLING 2014

# 25th International Conference on Computational Linguistics

# Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex-IV)

Workshop Chairs:
Michael Zock, Reinhard Rapp, Chu-Ren Huang

August 23, 2014
Dublin, Ireland

ISBN 978-1-873769-33-1
Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon
(CogALex-IV)
Michael Zock, Reinhard Rapp and Chu-Ren Huang (eds.)

# Introduction to the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex-IV)

## 1 Background[1]

Starting with a workshop devoted to electronic dictionaires at COLING, Geneva, 2004 (*Enhancing and Using Electronic Dictionaries*) we have continued to do so, by keeping it associated to this conference (2008, Manchester, 2010, Beijing , 2012, Mumbai).[2] What we did change though is the name, as CogALex captures better our mindset, i.e. the focus from which we look at the lexicon. CogALex stands for Cognitive Aspects of the Lexicon. Encouraged by the enthusiasm and interest expressed by the participants of the preceding events, it was natural to come up with a follow-up workshop.

As in the past our aim is to provide a forum for computational lexicographers, researchers in NLP, psychologists and users of lexical resources to share their knowledge and needs concerning the construction, organisation and use of a lexicon by people (lexical access) and machines (NLP, IR, datamining). Hence we invite again researchers with diverse backgrounds to address unsolved problems (see below). In addition we have added two features: One is devoted to a *shared Task*[3] where the sys-tem is meant to learn automatically how to find a word on the basis of its associated terms, the other feature is *vocabulary learning*.

While vocabulary learning is understandably not a hot topic in computational linguistics, it is nevertheless an important aspect of language learning, both in the mother tongue and in a foreign language. Having a rich stock of vocabulary is certainly an asset, but knowing the basic words and expressions is a must. Yet, people tend to forget some of the words they (thought to) have learned. This is not just a question of exercise (quantity vs. quality). Traditional word lists or flash cards are clearly not the ultimate answer. They are not only boring, but also rarely very effective. There is so much more we could do these days by using corpora and computational linguistics know-how, to extract the to-be learned words from text and to display them with their context. Hence, rather than having the user repeat sin-gle words (or word pairs) we could display them in various contexts (e.g. sentences), thereby making sure that the chosen ones correspond to the learners' level and interests.

## 2 Motivation

The way we look at dictionaries (their creation and use) has changed dramatically over the past 30 years. While being considered as an appendix to grammar in the past, by now they have moved to centre stage. Indeed, there is hardly any task in NLP which can be conducted without them. Also, rather than being static entities (database view), dictionaries are now viewed as dynamic networks, i.e. graphs, whose nodes and links (connection strengths) may change over time. Interestingly, properties concerning topology, clustering and evolution known from other disciplines (society, economy, human brain) also apply to dictionaries: everything is linked, hence accessible, and everything is evolving. Given these similarities, one may wonder what we can learn from these disciplines. In this 4th edition of the CogALex workshop we therefore also invited scientists working in these fields, with the goal to broaden the picture, i.e. to gain a better understanding concerning the mental lexicon and to integrate these findings into our dictionaries in order to support navigation. Given recent advances in neurosciences, it appears timely to seek inspiration from neuroscientists studying the human brain. There is also a lot to be learned from other fields studying graphs and networks, even if their object of study is something else than language, for example biology, economy or society.



[2]Workshop proceedings: see ACL Anthology: http://aclweb.org/anthology/

[3]http://pageperso.lif.univ-mrs.fr/∼michael.zock/CogALex-IV/cogalex-webpage/pst.html

## 3 Topics of Interest

This workshop is about possible enhancements of lexical resources and electronic dictionaries. To perform the groundwork for the next generation of such resources we invite researchers involved in the building of such tools. The idea is to discuss modifications of existing resources by taking the usersí needs and knowledge states into account, and to capitalize on the advantages of the digital media. For this workshop we solicit papers including but not limited to the following topics, each of which can be considered from various points of view: linguistics, neuro- or psycholinguistics (tip of the tongue problem, associations), network related sciences (sociology, economy, biology), mathematics (vector-based approaches, graph theory, small-world problem), etc.

I) Analysis of the conceptual input of a dictionary user

- What does a language producer start from (bag of words)?
- What is in the authors' minds when they are generating a message and looking for a word?
- What does it take to bridge the gap between this input and the desired output (target word)?

II) The meaning of words

- Lexical representation (holistic, decomposed);
- Meaning representation (concept based, primitives);
- Revelation of hidden information (distributional semantics, latent semantics, vector-based approaches: LSA/HAL);
- Neural models, neurosemantics, neurocomputational theories of content representation.

III) Structure of the lexicon

- Discovering structures in the lexicon: formal and semantic point of view (clustering, topical structure);
- Creative ways of getting access to and using word associations (reading between the lines, subliminal communication);
- Evolution, i.e. dynamic aspects of the lexicon (changes of weights);
- Neural models of the mental lexicon (distribution of information concerning words, organi-sation of words).

IV) Methods for crafting dictionaries or Indexes

- Manual, automatic or collaborative building of dictionaries and indexes (crowd-sourcing, serious games, etc.);
- Impact and use of social networks (Facebook, Twitter) for building dictionaries, for organizing and indexing the data (clustering of words), and for allowing to track navigational strategies, etc.;
- (Semi-) automatic induction of the link type (e.g. synonym, hypernym, meronym, association, collocation, ...);
- Use of corpora and patterns (datamining) for getting access to words, their uses, combinations and associations.

V) Dictionary access (navigation and search strategies) and interface issues

- Search based on sound, meaning or associations;

- Search (simple query vs multiple words);
- Context-dependent search (modification of usersí goals during search);
- Recovery;
- Navigation (frequent navigational patterns or search strategies used by people);
- Interface problems, data-visualisation.

We received 30 submissions, of which seven were accepted as full papers, eight were accepted for poster presentation, and nine were accepted in the context of the shared task. While we did not get papers on all the issues mentioned in our call, we did get a quite rich panel of topics including cognitive approaches to lexical access, considerations on word meaning and ontologies, manual and automatic approaches for constructing lexicons, as well as pragmatic aspects. It was also interesting to see the variety of languages in which these issues are addressed. In sum, the community working on dictionaries is dynamic, and there seems to be a growing awareness of the importance of some of the problems presented in our call for papers.

We would like to thank Roberto Navigli for having accepted to be our invited speaker, Shishang Wang for proofreading, and the COLING organizers for providing the framework and for their support. We would also like to express our sincerest thanks to all the members of the Programme Committee whose expertise was invaluable to assure 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 improve their work. In the hope that the results will inspire you, provoke fruitful discussions and result in future collaborations.

Dublin, Ireland, August 2014

Michael Zock, Reinhard Rapp, Chu-Ren Huang

**Organizers:**

Michael Zock (LIF-CNRS, Marseille, France)
Reinhard Rapp (LIF, Marseille, France and University of Mainz, Germany)
Chu-Ren Huang (The Hong Kong Polytechnic University, China)

**Invited Speaker:**

Roberto Navigli (Sapienza, University of Rome, Italy)

**Program Committee:**

Bel Enguix, Gemma (LIF Marseille, France)
Chang, Jason (National Tsing Hua University, Taiwan)
Cook, Paul (University of Melbourne, Australia)
Cristea, Dan (University A.I.Cuza, Iasi, Romania)
De Deyne, Simon (Experimental Psychology, Leuven, Belgium and Adelaide, Australia)
De Melo, Gerard (IIIS, Tsinghua University, Beijing, China)
Ferret, Olivier (CEA LIST, Gif sur Yvette, France)
Fontenelle, Thierry (CDT, Luxemburg)
Gala, Nuria (LIF-CNRS, Aix Marseille University, Marseille, France)
Granger, Sylviane (Université Catholique de Louvain, Belgium)
Grefenstette, Gregory (Inria, Saclay, France)
Hirst, Graeme (University of Toronto, Canada)
Hovy, Eduard (CMU, Pittsburgh, USA)
Hsieh, Shu-Kai (National Taiwan University, Taipei, Taiwan)
Huang, Chu-Ren (Hongkong Polytechnic University, China)
Joyce, Terry (Tama University, Kanagawa-ken, Japan)
Lapalme, Guy (RALI, University of Montreal, Canada)
Lenci, Alessandro (CNR, University of Pisa, Italy)
L'Homme, Marie Claude (University of Montreal, Canada)
Mihalcea, Rada (University of Texas, USA)
Navigli, Roberto (Sapienza, University of Rome, Italy)
Pirrelli, Vito (ILC, Pisa, Italy)
Polguère, Alain (ATILF-CNRS, Nancy, France)
Rapp, Reinhard (LIF Marseille, France and University of Mainz, Germany)
Rosso, Paolo (NLEL, Universitat Politècnica de València, Spain)
Schwab, Didier (LIG-GETALP, Grenoble, France)
Serasset, Gilles (IMAG, Grenoble, France)
Sharoff, Serge (University of Leeds, UK)
Su, Jun-Ming (University of Tainan, Taiwan)
Tiberius, Carole (Institute for Dutch Lexicology, The Netherlands)
Tokunaga, Takenobu (TITECH, Tokyo, Japan)
Tufis, Dan (RACAI, Bucharest, Romania)
Valitutti, Alessandro (Helsinki Institute of Information Technology, Finland)
Wandmacher, Tonio (IRT SystemX, Saclay, France)
Zock, Michael (LIF-CNRS, Marseille, France)

# Table of Contents

# Workshop Program

**Saturday, August 23, 2014**

9:00–9:05 **Opening Remarks**

9:05–10:30 **Session 1: Shared Task on the Lexical Access Problem**

*The CogALex-IV Shared Task on the Lexical Access Problem*
Reinhard Rapp and Michael Zock

*A Two-Stage Approach for Computing Associative Responses to a Set of Stimulus Words*
Urmi Ghosh, Sambhav Jain and Paul Soma

*Deep Learning from Web-Scale Corpora for Better Dictionary Interfaces*
Pavel Smrz and Lubomir Otrusina

*Exploring the use of word embeddings and random walks on Wikipedia for the CogAlex shared task*
Josu Goikoetxea, Eneko Agirre and Aitor Soroa

*ETS Lexical Associations System for the COGALEX-4 Shared Task*
Michael Flor and Beata Beigman Klebanov

*Using Significant Word Co-occurences for the Lexical Access Problem*
Rico Feist, Daniel Gerighausen, Manuel Konrad, Georg Richter, Thomas Eckart, Dirk Goldhahn and Uwe Quasthoff

*NaDiR: Naive Distributional Response Generation*
Gabriella Lapesa and Stefan Evert

*Retrieving Word Associations with a Simple Neighborhood Algorithm in a Graph-based Resource*
Gemma Bel Enguix

*Predicting sense convergence with distributional semantics: an application to the CogaLex 2014 shared task*
Laurianne Sitbon and Lance De Vine

*WordFinder*
Catalin Mititelu and Verginica Barbu Mititelu

**Saturday, August 23, 2014 (continued)**

10:30–11:00 **Coffee Break**

11:00–12:00 **Session 2: Invited Talk**

*(Digital) Goodies from the ERC Wishing Well: BabelNet, Babelfy, Video Games with a Purpose and the Wikipedia Bitaxonomy*
Roberto Navigli

12:00–12:30 **Session 3a: Booster Presentations for Posters**

12:03–12:06 *Measuring Similarity from Word Pair Matrices with Syntagmatic and Paradigmatic Associations*
Jin Matsuoka and Yves Lepage

12:06–12:09 *Jibiki-LINKS: a tool between traditional dictionaries and lexical networks for modelling lexical resources*
Ying Zhang, Mathieu Mangeot, Valérie Bellynck and Christian Boitet

12:09–12:12 *When Frequency Data Meet Dispersion Data in the Extraction of Multi-word Units from a Corpus: A Study of Trigrams in Chinese*
Chan-Chia Hsu

12:12–12:15 *Exploring Mental Lexicon in an Efficient and Economic Way: Crowdsourcing Method for Linguistic Experiments*
Shichang Wang, Chu-Ren Huang, Yao Yao and Angel Chan

12:15–12:18 *A Computational Approach to Generate a Sensorial Lexicon*
Serra Sinem Tekiroglu, Gözde Özbal and Carlo Strapparava

12:18–12:21 *Database Design of an Online E-Learning Tool of Chinese Classifiers*
Helena Gao

12:21–12:24 *Default Physical Measurements in SUMO*
Francesca Quattri, Pease Adam and John P. McCrae

12:24–12:27 *Lexical Access Preference and Constraint Strategies for Improving Multiword Expression Association within Semantic MT Evaluation*
Dekai Wu, Chi-kiu Lo and Markus Saers

12:30–14:00 **Lunch Break**

14:00–15:00 **Session 3b: Posters**

**Saturday, August 23, 2014 (continued)**

15:00–15:30 **Coffee Break**

15:30–17:40 **Session 4: Oral Presentations**

15:30–15:50 *A Lexical Network with a Morphological Model in It*
Nabil Gader, Aurore Koehl and Alain Polguère

15:50–16:10 *Dimensions of Metaphorical Meaning*
Andrew Gargett, Josef Ruppenhofer and John Barnden

16:10–16:30 *Constructing an Ontology of Japanese Lexical Properties: Specifying its Property Structures and Lexical Entries*
Terry Joyce and Bor Hodošček

16:30–16:50 *Frames and terminology: representing predicative terms in the field of the environment*
Marie-Claude L' Homme and Benoît Robichaud

16:50–17:10 *Modelling the Semantics of Adjectives in the Ontology-Lexicon Interface*
John P. McCrae, Francesca Quattri, Christina Unger and Philipp Cimiano

17:10–17:30 *Discovering Conceptual Metaphors using Source Domain Spaces*
Samira Shaikh, Tomek Strzalkowski, Kit Cho, Ting Liu, George Aaron Broadwell, Laurie Feldman, Sarah Taylor, Boris Yamrom, Ching-Sheng Lin, Ning Sa, Ignacio Cases, Yuliya Peshkova and Kyle Elliot

17:30–17:50 *Wordfinding Problems and How to Overcome them Ultimately With the Help of a Computer*
Michael Zock

17:50–18:00 **Conclusions and Closing**