

**\*\* LAST CALL FOR PAPERS \*\***

**EXTENDED DEADLINE:**

**New Paper Submission Deadline: October 2nd**

**NEW NOTIFICATION DATE: October 21st**

**COGALEx -V**

## **Cognitive Aspects of the Lexicon**

<https://sites.google.com/site/cogalex2016/home>

Workshop co-lated with COLING (the 26th International Conference on Computational Linguistics, Osaka, Japan), December 12, 2016

Invited speaker: *Chris Biemann* (Language Technology group, Universität Hamburg)

We are pleased to announce the 5th Workshop on ‘*Cognitive Aspects of the Lexicon*’ (COGALEx-V), taking place just before COLING (Osaka, Japan), December 12, 2016.

### **1 Context and background**

The way we look at the *lexicon* (creation and use) has changed dramatically over the past 30 years. While in the past being considered as an appendix to *grammar*, the lexicon has now moved to centre stage. Indeed, there is hardly any task in NLP which can be conducted without it. Also, rather than considering it as a static entity (database view), dictionaries are now viewed as dynamic networks, akin to the human brain, whose nodes and links (connection strengths) may change over time.

Linguists work on *products*, while psychologists and computer scientists deal with *processes*. They decompose the task into a set of subtasks, i.e. modules between which information flows. There are inputs, outputs and processes in between. A typical task in language processing is to go from meanings to sound or vice versa, the two extremes of language production and language understanding. Since this mapping is hardly ever direct, various intermediate steps or layers (syntax, morphology) are necessary.

Most of the work done by psycholinguists has dealt with the information flow from *meaning* (or concepts) to *sound* or the other way around. What has not been addressed though is the creation of a map of the *mental lexicon*, that is a representation of the way how words are organized or connected. In this respect WordNet and Roget's Thesaurus are probably the best one can expect these days. This being said, to find a word in a resource one has to reduce the search space (entire lexicon) and this is done via the knowledge one has at the onset of search. While the information stored in the lexicon is a *product*, its access is clearly a *process*, or more precisely, a cognitive, i.e. knowledge-based process.

#### **1.1 Goal**

The goal of COGALEx is to provide a forum for researchers in NLP, psychologists, computational lexicographers and users of lexical resources to share their knowledge and needs concerning the construction, organization and use of a lexicon by people (lexical access) and machines (NLP, IR, data-mining).

Like in the past (2004, 2008, 2010, 2012 and 2014), we will invite researchers to address various unsolved problems, by putting this time stronger emphasis though on *distributional semantics* (DS). Indeed, we would like to see work showing the relevance of DS as a cognitive model of the lexicon. The interest in distributional approaches has grown considerably over the last few year, both in computational linguistics and cognitive sciences. A further boost has been provided by the recent hype around *deep learning* and *neural embeddings*. While all these approaches seem to have great potential, their added value to address cognitive and semantic aspects of the lexicon still needs to be shown.

This workshop is about possible enhancements of lexical resources and electronic dictionaries, as well as on any aspect relevant to the achieve a better understanding of the mental lexicon and semantic memory. We solicit contributions including but not limited to the topics listed here below, topics, which can be considered from any of the following points of view:

- (computational, corpus) 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.

We also plan to organize a “friendly competition” for corpus-based models of lexical networks and navigation, i.e. lexical access (see below).

## 1.2 Possible Topics

### 1.2.1 Analysis of the conceptual input of a dictionary user

- What does a language producer start out with and how does this input relate to the target form? (meaning, collocation, topically related, etc.)
- 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)?

### 1.2.2 The meaning of words

- Lexical representation (holistic, decomposed)
- Meaning representation (concept based, primitives)
- Distributional semantics (count models, neural embeddings, etc. )
- Neurocomputational theories of content representation.

### 1.2.3 Structure of the lexicon

- Discovering structures in the lexicon: formal and semantic point of view (clustering, topical structure)
- Evolution, i.e. dynamic aspects of the lexicon (changes of weights)
- Neural models of the mental lexicon (distribution of information concerning words, organization of words)

### 1.2.4 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 (data-mining) for getting access to words, their uses, combinations and associations

### 1.2.5 Dictionary access (navigation and search strategies), interface issues,

- Search based on sound, meaning or associations

- Search (simple query vs. multiple words)
- Search-space determination based on user's knowledge, meta-knowledge and cognitive state (information available at the onset, knowledge concerning the relationship between the input and the target word, ...)
- Context-dependent search (modification of users' goals during search)
- Navigation (frequent navigational patterns or search strategies used by people)
- Interface problems, data-visualization
- Creative ways of getting access to and using word associations (reading between the lines, subliminal communication).

## 2 Description of the shared tasks associated with the workshop.

As part of the workshop, we propose a shared task concerning the corpus-based identification of semantic relations. The goal of this "competition between gentlemen" is less the discovery of the best system, as the testing of the relative efficiency of different distributional models and other corpus-based approaches on a challenging semantic task. We will provide the training and test data, and the participants are expected to submit a short paper (4 pages) describing their approach and evaluation results (using the official scoring scripts), together with the output produced by their system on the test data.

For more details see : <https://sites.google.com/site/cogalex2016/home/shared-task>

## 3 INVITED SPEAKER

*Chris Biemann*, well known (among other things) for his work on [graph-based NLP](#), has kindly accepted to give the invited presentation. Leader of the LT research group (TU Darmstadt), Chris is now affiliated with the *Language Technology* group of the University of Hamburg.

## 4 Deadlines.

- Paper **submission deadline** : October 2<sup>nd</sup>, 2016
- Author **notification** : October 21st, 2016
- Camera ready : October 30, 2016
- Workshop : December 12

For shared task

- September 26: Expression of interest (e-mail to [esantus@gmail.com](mailto:esantus@gmail.com))
- October 15: **Submission** of *system description* papers (4+1 pages) and *system output*
- October 25: Reviews returned
- October 30: Camera ready due by Authors

## 5 Submission

The submissions should be written in English and be anonymized for review. They must comply with the style-sheets provided by Coling: <http://coling2016.anlp.jp/#instructions>

- Long papers may consist of 8 pages of content, plus 2 pages for references;
- Short paper may consist of up to 4 pages of content, plus 2 pages for references
- The respective final versions may be up to 9 pages for long papers and 5 pages for short ones. In both cases the number of pages for references is limited to 3 pages.

Papers should be in PDF format and have to be submitted electronically via the START submission system (<https://www.softconf.com/coling2016/CogALex-V/>). You probably have to register first,

and then choose: submission, i.e. (<https://www.softconf.com/coling2016/CogALex-V/user/scmd.cgi?scmd=submitPaperCustom&pageid=0>).

## 6 Organizers.

- Michael Zock (LIF, CNRS, Aix-Marseille University, Marseille, France)
- Alessandro Lenci (Computational Linguistics Laboratory, University of Pisa,, Italy)
- Stefan Evert (FAU, Erlangen-Nürnberg, Germany)

## 7 Contact persons

For general questions, please get in touch with *Michael Zock* ([michael.zock@lif.univ-mrs.fr](mailto:michael.zock@lif.univ-mrs.fr)), for questions concerning the shared task, send an e-mail to *Stefan Evert* ([stefan.evert@fau.de](mailto:stefan.evert@fau.de)).

## 8 Program committee

- Biemann, Chris (Universität Hamburg, Germany)
- Babych, Bogdan (University of Leeds, UK)
- Brysbaert, Marc (Experimental Psychology, Ghent University, Belgium)
- Cristea, Dan ("Al. I. Cuza" University, Iasi, Romania)
- deDeyne, Simon (University of Adelaide, Australia)
- de Melo, Gerard (IIS, Tsinghua University, Beijing, China)
- Evert, Stefan (University of Erlangen, Germany)
- Ferret, Olivier (CEA LIST, France)
- Fontenelle, Thierry (CDT, Luxembourg)
- Gala, Nuria (University of Aix-Marseille, France)
- Geeraerts, Dirk (University of Leuven, Belgium)
- Granger, Sylviane (Université Catholique de Louvain, Belgium)
- Grefenstette, Gregory (Inria, Paris, France)
- Hirst, Graeme (University of Toronto, Canada)
- Hovy, Ed (CMU, Pittsburgh, USA)
- Hsieh, Shu-Kai (National Taiwan University, Taipei, Taiwan)
- Joyce, Terry (Tama University, Kanagawa-ken, Japan)
- Lafourcade, Matthieu (LIRMM, université de Montpellier, France)
- Lapalme, Guy (RALI, University of Montreal, Canada)
- Lebani, Gianluca (University of Pisa, Italy)
- Lenci, Alessandro (University of Pisa, Italy)
- L'Homme, Marie Claude (University of Montreal, Canada)
- Mititelu, Verginica (RACAI, Bucharest, Romania)
- Navigli, Roberto (Sapienza, Rome, Italy)
- Paradis, Carita (Centre for Languages and Literature Lund University, Sweden)
- Pihlevar, Taher (university of Cambridge, UK)
- Pirrelli, Vito (ILC, Pisa, Italy)
- Polguère, Alain (ATILF-CNRS, Nancy, France)
- Purver, Matthew (King's College, London, UK)
- Ramisch, Carlos (AMU, Marseille, France)
- Rayson, Paul (UCREL, university of Lancaster, UK)
- Rosso, Paol (NLEL, Universitat Politècnica de València, Spain)
- Sahlgren, Magnus (Gavagai Inc. & SICS, Sweden)
- Schulte im Walde, Sabine (University of Stuttgart, Germany)
- Schwab, Didier (LIG, Grenoble, France)
- Sharoff, Serge (University of Leeds, UK)
- Stella, Massimo (Institute for Complex Systems Simulation, university of Southampton, UK)
- Tokunaga, Takenobu (TITECH, Tokyo, Japan)
- Tufis, Dan (RACAI, Bucharest, Romania)

- Zarcone, Alessandra (Saarland University, Germany)
- Zock, Michael (LIF-CNRS, Marseille, France)