

EMNLP 2015

**CONFERENCE ON
EMPIRICAL METHODS IN
NATURAL LANGUAGE PROCESSING**

Proceedings of the Sixth Workshop on Cognitive Aspects of
Computational Language Learning
(CogACLL-2015)

18 September 2015
Lisbon, Portugal

Order print-on-demand copies from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571 USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

©2015 The Association for Computational Linguistics
ISBN: 978-1-941643-32-7

Preface

The Workshop on Cognitive Aspects of Computational Language Learning (CogACLL) took place on September 18, 2015 in Lisbon, Portugal, in conjunction with the EMNLP 2015. The workshop was endorsed by ACL Special Interest Group on Natural Language Learning (SIGNLL). This is the sixth edition of related workshops first held with ACL 2007, EACL 2009, 2012 and 2014 and as a standalone event in 2013.

The workshop is targeted at anyone interested in the relevance of computational techniques for understanding first, second and bilingual language acquisition and change or loss in normal and pathological conditions.

The human ability to acquire and process language has long attracted interest and generated much debate due to the apparent ease with which such a complex and dynamic system is learnt and used on the face of ambiguity, noise and uncertainty. This subject raises many questions ranging from the nature vs. nurture debate of how much needs to be innate and how much needs to be learned for acquisition to be successful, to the mechanisms involved in this process (general vs specific) and their representations in the human brain. There are also developmental issues related to the different stages consistently found during acquisition (e.g. one word vs. two words) and possible organizations of this knowledge. These have been discussed in the context of first and second language acquisition and bilingualism, with cross linguistic studies shedding light on the influence of the language and the environment.

The past decades have seen a massive expansion in the application of statistical and machine learning methods to natural language processing (NLP). This work has yielded impressive results in numerous speech and language processing tasks, including e.g. speech recognition, morphological analysis, parsing, lexical acquisition, semantic interpretation, and dialogue management. The good results have generally been viewed as engineering achievements. Recently researchers have begun to investigate the relevance of computational learning methods for research on human language acquisition and change. The use of computational modeling is a relatively recent trend boosted by advances in machine learning techniques, and the availability of resources like corpora of child and child-directed sentences, and data from psycholinguistic tasks by normal and pathological groups. Many of the existing computational models attempt to study language tasks under cognitively plausible criteria (such as memory and processing limitations that humans face), and to explain the developmental stages observed in the acquisition and evolution of the language abilities. In doing so, computational modeling provides insight into the plausible mechanisms involved in human language processes, and inspires the development of better language models and techniques. These investigations are very important since if computational techniques can be used to improve our understanding of human language acquisition and change, these will not only benefit cognitive sciences in general but will reflect back to NLP and place us in a better position to develop useful language models.

We invited submissions on relevant topics, including:

- Computational learning theory and analysis of language learning and organization
- Computational models of first, second and bilingual language acquisition

- Computational models of language changes in clinical conditions
- Computational models and analysis of factors that influence language acquisition and use in different age groups and cultures
- Computational models of various aspects of language and their interaction effect in acquisition, processing and change
- Computational models of the evolution of language
- Data resources and tools for investigating computational models of human language processes
- Empirical and theoretical comparisons of the learning environment and its impact on language processes
- Cognitively oriented Bayesian models of language processes
- Computational methods for acquiring various linguistic information (related to e.g. speech, morphology, lexicon, syntax, semantics, and discourse) and their relevance to research on human language acquisition
- Investigations and comparisons of supervised, unsupervised and weakly-supervised methods for learning (e.g. machine learning, statistical, symbolic, biologically-inspired, active learning, various hybrid models) from a cognitive perspective.

Acknowledgements

We would like to thank the members of the Program Committee for the timely reviews and the authors for their valuable contributions. Aline Villavicencio is partly funded by projects CNPq 551964/2011-1, 312184/2012-3 and 482520/2012-4 and Samsung/UFRGS 4287, Alessandro Lenci by project CombiNet (PRIN 2010-11 20105B3HE8) funded by the Italian Ministry of Education, University and Research (MIUR).

Robert Berwick
 Anna Korhonen
 Alessandro Lenci
 Thierry Poibeau
 Aline Villavicencio

Organizers:

Robert Berwick (Massachusetts Institute of Technology, USA)
Anna Korhonen (University of Cambridge, UK)
Alessandro Lenci (University of Pisa, Italy)
Thierry Poibeau (LATTICE-CNRS, France)
Aline Villavicencio (Federal University of Rio Grande do Sul, Brazil)

Program Committee:

Afra Alishahi, Tilburg University (Netherlands)
Colin J Bannard, University of Texas at Austin (USA)
Philippe Blache, LPL-CNRS (France)
Susana Bautista Blasco, Federal University of Rio Grande do Sul (Brazil)
Antal van den Bosch, Radboud University Nijmegen (Netherlands)
Ted Briscoe, University of Cambridge (UK)
Grzegorz Chrupała, Tilburg University (Netherlands)
Robin Clark, University of Pennsylvania (USA)
Matthew W. Crocker, Saarland University (Germany)
Walter Daelemans, University of Antwerp (Belgium)
Dan Dediu, Max Planck Institute for Psycholinguistics (The Netherlands)
Barry Devereux, University of Cambridge (UK)
Ted Gibson, Massachusetts Institute of Technology (USA)
Sharon Goldwater, University of Edinburgh (UK)
Marco Idiart, Federal University of Rio Grande do Sul (Brazil)
Mark Johnson, Macquarie University (Australia)
Aravind Joshi, University of Pennsylvania (USA)
Gianluca Leboni, University of Pisa (Italy)
Igor Malioutov, Massachusetts Institute of Technology (USA)
Marie-Catherine de Marneffe, The Ohio State University (USA)
Brian Murphy, Queen's University Belfast (UK)
Tim O'Donnell, Massachusetts Institute of Technology (USA)
Muntsa Padró, Nuance (Canada)
Lisa Pearl, University of California - Irvine (USA)
Massimo Poesio, University of Trento (Italy)
Ari Rappoport, The Hebrew University of Jerusalem (Israel)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Ekaterina Shutova, University of California, Berkeley (USA)
Maity Siqueira, Federal University of Rio Grande do Sul (Brazil)
Mark Steedman, University of Edinburgh (UK)
Remi van Trijp, Sony Computer Science Laboratory Paris (France)
Charles Yang, University of Pennsylvania (USA)

Menno van Zaanen, Tilburg University (Netherlands)
Alessandra Zarcone, Saarland University (Germany)
Leonardo Zilio, Federal University of Rio Grande do Sul (Brazil)

Table of Contents

<i>Using reading behavior to predict grammatical functions</i> Maria Barrett and Anders Sjøgaard	1
<i>Reading metrics for estimating task efficiency with MT output</i> Sigrid Klerke, Sheila Castilho, Maria Barrett and Anders Sjøgaard	6
<i>Evaluating Models of Computation and Storage in Human Sentence Processing</i> Thang Luong, Timothy O'Donnell and Noah Goodman	14
<i>An agent-based model of a historical word order change</i> Jelke Bloem, Arjen Versloot and Fred Weerman	22
<i>Towards a Model of Prediction-based Syntactic Category Acquisition: First Steps with Word Embeddings</i> Robert Grimm, Giovanni Cassani, Walter Daelemans and Steven Gillis	28
<i>Which distributional cues help the most? Unsupervised contexts selection for lexical category acquisition</i> Giovanni Cassani, Robert Grimm, Walter Daelemans and Steven Gillis	33
<i>Language Emergence in a Population of Artificial Agents Equipped with the Autotelic Principle</i> Miquel Cornudella and Thierry Poibeau	40
<i>A Computational Study of Cross-situational Lexical Learning of Brazilian Portuguese</i> Pablo Faria	45
<i>Units in segmentation: a computational investigation</i> Çağrı Çöltekin	55
<i>Estimating Grammatical Redundancy by Measuring Their Importance for Syntactic Parser Performance</i> Aleksandrs Berdicevskis	65
<i>Improving Coordination on Novel Meaning through Context and Semantic Structure</i> Thomas Brochhagen	74
<i>Perceptual, conceptual, and frequency effects on error patterns in English color term acquisition</i> Barend Beekhuizen and Suzanne Stevenson	83
<i>Motif discovery in infant- and adult-directed speech</i> Bogdan Ludusan, Amanda Seidl, Emmanuel Dupoux and Alex Cristia	93
<i>Modeling dative alternations of individual children</i> Antal van den Bosch and Joan Bresnan	103

Workshop Program

Friday, September 18, 2015

09:00–09:10 **Opening and Introduction**

09:10–10:30 **Session 1: Language Processing**

09:10–09:30 *Using reading behavior to predict grammatical functions*
Maria Barrett and Anders Søgaard

09:30–10:00 *Reading metrics for estimating task efficiency with MT output*
Sigrid Klerke, Sheila Castilho, Maria Barrett and Anders Søgaard

10:00–10:30 *Evaluating Models of Computation and Storage in Human Sentence Processing*
Thang Luong, Timothy O'Donnell and Noah Goodman

10:30–11:00 *Coffee Break*

11:00–11:50 **Invited Talk by Afra Alishahi**

11:50–12:10 **Session 2: Language Change**

11:50–12:10 *An agent-based model of a historical word order change*
Jelke Bloem, Arjen Versloot and Fred Weerman

Friday, September 18, 2015 (continued)

12:10–13:00 Session 3: Poster Session

12:10–13:00 *Towards a Model of Prediction-based Syntactic Category Acquisition: First Steps with Word Embeddings*

Robert Grimm, Giovanni Cassani, Walter Daelemans and Steven Gillis

12:10–12:30 *Which distributional cues help the most? Unsupervised contexts selection for lexical category acquisition*

Giovanni Cassani, Robert Grimm, Walter Daelemans and Steven Gillis

12:10–13:00 *Language Emergence in a Population of Artificial Agents Equipped with the Autotelic Principle*

Miquel Cornudella and Thierry Poibeau

12:10–13:00 *A Computational Study of Cross-situational Lexical Learning of Brazilian Portuguese*

Pablo Faria

12:10–13:00 *Units in segmentation: a computational investigation*

Çağrı Çöltekin

12:10–13:00 *Estimating Grammeme Redundancy by Measuring Their Importance for Syntactic Parser Performance*

Aleksandrs Berdicevskis

13:00–14:10 Lunch

14:10–15:00 Invited Talk by Antal van den Bosch

Friday, September 18, 2015 (continued)

15:00–15:30 Session 4: Language Processing II

15:00–15:30 *Improving Coordination on Novel Meaning through Context and Semantic Structure*
Thomas Brochhagen

15:30–16:00 Coffee Break

16:00–17:30 Session 5: Language Acquisition

16:00–16:30 *Perceptual, conceptual, and frequency effects on error patterns in English color term acquisition*
Barend Beekhuizen and Suzanne Stevenson

16:30–17:00 *Motif discovery in infant- and adult-directed speech*
Bogdan Ludusan, Amanda Seidl, Emmanuel Dupoux and Alex Cristia

17:00–17:30 *Modeling dative alternations of individual children*
Antal van den Bosch and Joan Bresnan

17:30–17:35 Closing Session

