

EACL 2014

**14th Conference of the European Chapter
of the Association for Computational Linguistics**



**Proceedings of the Workshop on Cognitive Aspects of
Computational Language Learning
(CogACLL)**

April 26 2014
Gothenburg, Sweden

©2014 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-84-8

Introduction

The Workshop on Cognitive Aspects of Computational Language Learning (CogACLL) took place on April 26, 2014 in Gothenburg, Sweden, in conjunction with the 14th Conference of the European Chapter of the Association for Computational Linguistics. The workshop was endorsed by ACL Special Interest Group on Natural Language Learning (SIGNLL). This is the fifth edition of related workshops that was first held at ACL 2007 in Prague, EACL 2009 in Athens, EACL 2012 in Avignon and as a standalone event in Paris 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.

Submissions included works on specific languages like English, Portuguese and German, along with crosslinguistic studies. Besides paper presentations the technical program included two invited talks by Philippe Blache, from Aix-Marseille Université and CNRS (France) and Alexander Clark, from King's College London (UK).

Acknowledgements

We would like to thank the members of the Program Committee for the timely reviews and the authors for their valuable contributions. Thierry Poibeau is partly funded by TransferS (laboratoire d'excellence, program "Investissements d'avenir" ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0099); Aline Villavicencio by projects CNPq 482520/2012-4, 478222/2011-4, 312184/2012-3 and 551964/2011-1; Muntsa Padró by project CAPES PNPd-2484/2009.

Alessandro Lenci
Muntsa Padró
Thierry Poibeau
Aline Villavicencio

Workshop Chairs:

Alessandro Lenci, University of Pisa (Italy)
Muntsa Padró, Federal University of Rio Grande do Sul (Brazil)
Thierry Poibeau, LATTICE-CNRS (France)
Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil)

Invited Speakers:

Philippe Blache, Aix-Marseille Université and CNRS (France)
Alexander Clark, King's College, London (UK)

Program Committee:

Afra Alishahi, Tilburg University (Netherlands)
Colin J Bannard, University of Texas at Austin (USA)
Marco Baroni, University of Trento (Italy)
Robert Berwick, Massachusetts Institute of Technology (USA)
Philippe Blache, Aix-Marseille Université and CNRS (France)
Jim Blevins, University of Cambridge (UK)
Antal van den Bosch, Radboud University Nijmegen (Netherlands)
Chris Brew, Nuance Communications (USA)
Ted Briscoe, University of Cambridge (UK)
Alexander Clark, Royal Holloway, University of London (UK)
Robin Clark, University of Pennsylvania (USA)
Stephen Clark, University of Cambridge (UK)
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)
Benjamin Fagard, Lattice-CNRS (France)
Jeroen Geertzen, University of Cambridge (UK)
Ted Gibson, Massachusetts Institute of Technology (USA)
Henriette Hendriks, University of Cambridge (UK)
Marco Idiart, Federal University of Rio Grande do Sul (Brazil)
Mark Johnson, Macquarie University (Australia)
Aravind Joshi, University of Pennsylvania (USA)
Gianluca Lebani, University of Pisa (Italy)
Igor Malioutov, Massachusetts Institute of Technology (USA)
Marie-Catherine de Marneffe, The Ohio State University (USA)
Maria Alice de Mattos Pimenta, Federal University of ABC (Brazil)
Maria Alice Parente, Federal University of Rio Grande do Sul (Brazil)
Massimo Poesio, University of Trento (Italy)
Brechtje Post, University of Cambridge (UK)
Ari Rappoport, The Hebrew University of Jerusalem (Israel)
Anne Reboul, L2C2-CNRS (France)
Kenji Sagae, University of Southern California (USA)
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)

Suzanne Stevenson, University of Toronto (Canada)
Remi van Trijp, Sony Computer Science Laboratory Paris (France)
Shuly Wintner, University of Haifa (Israel)
Charles Yang, University of Pennsylvania (USA)
Beracah Yankama, Massachusetts Institute of Technology (USA)
Menno van Zaanen, Tilburg University (Netherlands)
Alessandra Zarcone, University of Stuttgart (Germany)

Table of Contents

<i>Challenging incrementality in human language processing: two operations for a cognitive architecture</i> Philippe Blache	1
<i>A Brazilian Portuguese Phonological-prosodic Algorithm Applied to Language Acquisition: A Case Study</i> Vera Vasilévski, Márcio José Araujo and Helena Ferro Blasi	3
<i>Bayesian inference as a cross-linguistic word segmentation strategy: Always learning useful things</i> Lawrence Phillips and Lisa Pearl	9
<i>Learning the hyperparameters to learn morphology</i> Stella Frank	14
<i>An explicit statistical model of learning lexical segmentation using multiple cues</i> Çağrı Çöltekin and John Nerbonne	19
<i>Distributional Learning as a Theory of Language Acquisition</i> Alexander Clark	29
<i>A multimodal corpus for the evaluation of computational models for (grounded) language acquisition</i> Judith Gaspers, Maximilian Panzner, Andre Lemme, Philipp Cimiano, Katharina J. Rohlfing and Sebastian Wrede	30
<i>Towards a computational model of grammaticalization and lexical diversity</i> Christian Bentz and Paula Buttery	38
<i>How well can a corpus-derived co-occurrence network simulate human associative behavior?</i> Gemma Bel Enguix, Reinhard Rapp and Michael Zock	43
<i>Agent-based modeling of language evolution</i> Torvald Lekvam, Björn Gambäck and Lars Bungum	49
<i>Missing Generalizations: A Supervised Machine Learning Approach to L2 Written Production</i> Daniel Wiechmann and Elma Kerz	55

Conference Program

Saturday April 26, 2014

9:30 Opening and Introduction

Invited Talk 1

9:40 *Challenging incrementality in human language processing: two operations for a cognitive architecture*

Philippe Blache

10:30 Coffee Break

Session 1: Phonology, morphology and word segmentation

11:00 *A Brazilian Portuguese Phonological-prosodic Algorithm Applied to Language Acquisition: A Case Study*

Vera Vasilévski, Márcio José Araujo and Helena Ferro Blasi

11:20 *Bayesian inference as a cross-linguistic word segmentation strategy: Always learning useful things*

Lawrence Phillips and Lisa Pearl

11:40 *Learning the hyperparameters to learn morphology*

Stella Frank

12:00 *An explicit statistical model of learning lexical segmentation using multiple cues*

Çağrı Çöltekin and John Nerbonne

12:30 Lunch break

Saturday April 26, 2014 (continued)

Invited Talk 2

2:00 *Distributional Learning as a Theory of Language Acquisition*
Alexander Clark

Session 2: Lexical acquisition and language evolution

2:50 *A multimodal corpus for the evaluation of computational models for (grounded) language acquisition*

Judith Gaspers, Maximilian Panzner, Andre Lemme, Philipp Cimiano, Katharina J. Rohlfing and Sebastian Wrede

3:20 Coffee Break

3:45 *Towards a computational model of grammaticalization and lexical diversity*
Christian Bentz and Paula Buttery

4:05 *How well can a corpus-derived co-occurrence network simulate human associative behavior?*

Gemma Bel Enguix, Reinhard Rapp and Michael Zock

4:25 *Agent-based modeling of language evolution*

Torvald Lekvam, Björn Gambäck and Lars Bungum

Session 3: Second language acquisition

4:45 *Missing Generalizations: A Supervised Machine Learning Approach to L2 Written Production*

Daniel Wiechmann and Elma Kerz

5:15 Closing