EACL 2009

Proceedings of the EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition

31 March 2009
Megaron Athens International Conference Centre
Athens, Greece

Production and Manufacturing by
TEHNOGRAFIA DIGITAL PRESS
7 Ektoros Street
152 35 Vrilissia
Athens, Greece

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

Preface

This volume contains the papers accepted for presentation at the EACL 2009 Workshop on Cognitive Aspects of Computational Language Acquisition, held in Athens, Greece on the 31st of March, 2009. This workshop is the second of a series which was initiated during ACL 2007, held in Prague. The first edition of the workshop was organised by Anna Korhonen, Paula Buttery and Aline Villavicencio.

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 speech recognition, morphological analysis, parsing, lexical acquisition, semantic interpretation, and dialogue management. Advances in these areas are generally viewed as engineering achievements, but recently researchers have begun to investigate the relevance of computational learning techniques to research on human language acquisition. These investigations have double significance since an improved understanding of human language acquisition will not only benefit cognitive sciences in general, but may also feed back to the NLP community, placing researchers in a better position to develop new language models and/or techniques.

Success in this type of research requires close collaboration between NLP and cognitive scientists. The aim of this workshop is thus to bring together researchers from the diverse fields of NLP, machine learning, artificial intelligence, linguistics, psycho-linguistics, etc. who are interested in the relevance of computational techniques for understanding human language learning. The workshop is intended to bridge the gap between the computational and cognitive communities, promote knowledge and resource sharing, and help initiate interdisciplinary research projects.

In the call for papers we solicited papers describing cognitive aspects of computational language acquisition. The programme committee has selected 7 papers for publication that are representative of the state-of-the-art in this interdisciplinary area. Each full-length submission was independently reviewed by three members of the program committee, who then collectively faced the difficult task of selecting a subset of papers for publication from a very strong field.

We would like to thank our two invited speakers, Massimo Poesio and Robert Berwick, all the authors who submitted papers, as well as the members of the programme committee for the time and effort they contributed in reviewing the papers. Our thanks go also to the organisers of the main conference, the publication chairs, and the conference workshop committee headed by Miriam Butt and Stephen Clark.

Afra Alishahi, Thierry Poibeau and Aline Villavicencio

Organizers

Organizers:

Afra Alishahi, University of Saarland (Germany)
Thierry Poibeau, CNRS and University Paris 13 (France)
Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil) and University of Bath (UK)

Program Committee:

Colin J Bannard, Max Planck Institute for Evolutionary Anthropology (Germany)
Marco Baroni, University of Trento (Italy)
Robert C. Berwick, Massachusetts Institute of Technology (USA)
Jim Blevins, University of Cambridge (UK)
Rens Bod, University of Amsterdam (Netherlands)
Antal van den Bosch, Tilburg University (Netherlands)
Chris Brew, Ohio State University (USA)
Ted Briscoe, University of Cambridge (UK)
Robin Clark, University of Pennsylvania (USA)
Stephen Clark, University of Oxford (UK)
Matthew W. Crocker, Saarland University (Germany)
James Cussens, University of York (UK)
Walter Daelemans, University of Antwerp (Belgium) and Tilburg University (Netherlands)
Ted Gibson, Massachusetts Institute of Technology (USA)
Henriette Hendriks, University of Cambridge (UK)
Julia Hockenmaier, University of Illinois at Urbana-Champaign (USA)
Marco Idiart, Federal University of Rio Grande do Sul (Brazil)
Mark Johnson, Brown University (USA)
Aravind Joshi, University of Pennsylvania (USA)
Anna Korhonen, University of Cambridge (UK)
Alessandro Lenci, University of Pisa (Italy)
Massimo Poesio, University of Trento (Italy)
Brechtje Post, University of Cambridge (UK)
Ari Rappoport, The Hebrew University of Jerusalem (Israel)
Dan Roth, University of Illinois at Urbana-Champaign (USA)
Kenji Sagae, University of Southern California (USA)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Mark Steedman, University of Edinburgh (UK)
Suzanne Stevenson, University of Toronto (Canada)
Patrick Sturt, University of Edinburgh (UK)
Bert Vaux, University of Wisconsin (USA)
Charles Yang, University of Pennsylvania (USA)
Menno van Zaanen, Macquarie University (Australia)
Michael Zock, LIF, CNRS, Marseille (France)

Invited Speakers:

Robert Berwick, Massachusetts Institute of Technology (USA)
Massimo Poesio, University of Essex (UK) and University of Trento (Italy)

Table of Contents

Conference Program

Tuesday, March 31, 2009

Session 1: Introduction

9:20–9:30 Opening Remarks

9:30–10:30 Invited Talk by Massimo Poesio

10:30–11:00 Coffee Break

Session 2: Theoretical and Practical Aspects of Language Acquisition

11:00–11:30 *Towards a Formal View of Corrective Feedback*
Staffan Larsson and Robin Cooper

11:30–12:00 *A Collaborative Tool for the Computational Modelling of Child Language Acquisition*
Kris Jack

12:00–12:30 *What's in a Message?*
Stergos Afantenos and Nicolas Hernandez

12:30–14:00 Lunch Break

Session 3: Learnability and Grammatical Inference

14:00–15:00 Invited Talk by Robert Berwick

15:00–15:30 *Another Look at Indirect Negative Evidence*
Alexander Clark and Shalom Lappin

15:30–16:00 *Categorizing Local Contexts as a Step in Grammatical Category Induction*
Markus Dickinson and Charles Jochim

16:00–16:30 Coffee Break

Tuesday, March 31, 2009 (continued)

Session 4: The Ecology of Language

16:30–17:00 *Darwinised Data-Oriented Parsing - Statistical NLP with Added Sex and Death*
Dave Cochran

17:00–17:30 *Language Diversity across the Consonant Inventories: A Study in the Framework of Complex Networks*
Monojit Choudhury, Animesh Mukherjee, Anupam Basu, Niloy Ganguly, Ashish Garg and Vaibhav Jalan

17:30–17:45 Closing Remarks

Invited Talks

Conceptual Descriptions: Evidence from Corpora, the Mind, and the Brain
Massimo Poesio

All too often work in computational linguistics on the acquisition of conceptual descriptions takes place in isolation from work on concepts in psychology and neural science. We feel this is a mistake as evidence from these related disciplines can provide us with better ways of evaluating our results. In the talk I will present work in CIMEC on using cognitive evidence to evaluate the results of lexical acquisition work - specifically, using feature norms to evaluate the acquisition of features, and using EEG data to evaluate the results of categorization experiments.

Joint work with Marco Baroni, Brian Murphy, Eduard Barbu, and Abdulrahman Almuhareb, among others.

Treebank Parsing and Knowledge of Language: A Cognitive Perspective
Robert C. Berwick

Over the past 15 years, there has been increasing use of linguistically annotated sentence collections such as the LDC Penn Tree Bank (PTB) for constructing statistically based parsers. While these parsers have generally been built for engineering purposes, more recently such approaches have been advanced as potential cognitive solutions, e.g., for the problem of human language acquisition. Here we examine this possibility critically: we assess how well these Treebank parsers actually approach human/child language competence. We find that such systems fail to replicate many, perhaps most, empirically attested grammaticality judgments; seem overly sensitive, rather than robust, to training data idiosyncrasies; and easily acquire unnatural syntactic constructions never attested in human languages. Overall, we conclude that existing statistically based treebank parsers fail to incorporate much knowledge of language in these three senses. We discuss the implications of these results for the improvement of Treebank parsers and their cognitive relevance.

Joint work with Professor Sandiway Fong, University of Arizona.