ACL 2007

PRAGUE

ACL 2007

Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition

June 29, 2007
Prague, Czech Republic

The Association for Computational Linguistics

Production and Manufacturing by
Omnipress
2600 Anderson Street
Madison, WI 53704
USA

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 ACL 2007 Workshop on Cognitive Aspects of Computational Language Acquisition, held in Prague, Czech Republic on the 29th of June, 2007.

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 could 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 including:

- Computational learning theory and analysis of language learning
- Computational models of human (first, second and bilingual) language acquisition
- Computational models of various components of the language faculty and their impact on the acquisition task
- Computational models of the evolution of language
- Data resources and tools for investigating computational models of human language acquisition
- Empirical and theoretical comparisons of the learning environment and its impact on the acquisition task
- 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 the cognitive aspect

Of the 22 papers submitted, the programme committee selected 12 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. Among the accepted papers we see proposed techniques for creating, analysing and annotating data resources for research on language acquisition. We also see presentations of computational models for first and second language acquisition. These models investigate the acquisition of both syntactic and semantic phenomena, adopting different linguistic theories and formalisms, using varying levels of supervision.

We would like to thank 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 Simone Teufel.

Paula Buttery, Aline Villavicencio, Anna Korhonen

Organizers

Chairs:

Paula Buttery (University of Cambridge, UK)
Aline Villavicencio (Federal University of Rio Grande do Sul, Brazil, University of Bath, UK)
Anna Korhonen (University of Cambridge, UK)

Program Committee:

Colin J Bannard (Max Planck Institute for Evolutionary Anthropology, Germany)
Robert C. Berwick (Massachusetts Institute of Technology, USA)
Jim Blevins (University of Cambridge, UK)
Antal van den Bosch (Tilburg University, The Netherlands)
Chris Brew (Ohio State University, 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 Oxford, UK)
Matthew W. Crocker (Saarland University, Germany)
James Cussens (University of York, UK)
Walter Daelemans (University of Antwerp, Belgium and Tilburg University, The Netherlands)
Bruno Gaume (Universite Paul Sabatier, France)
Ted Gibson (Massachusetts Institute of Technology, USA)
Henriette Hendriks (University of Cambridge, UK)
Julia Hockenmaier (University of Pennsylvania, USA)
Marco Idiart (Federal University of Rio Grande do Sul, Brazil)
Mark Johnson (Brown University, USA)
Gea de Jong (University of Cambridge, UK)
Aravind Joshi (University of Pennsylvania, USA)
Gerard Kempen (Leiden University, Netherlands)
Brian MacWhinney (Carnegie Mellon University, USA)
Martin Pickering (University of Glasgow, UK)
Thierry Poibeau (University Paris 13, France)
Brechtje Post (University of Cambridge, UK)
Ari Rappoport (The Hebrew University of Jerusalem, Israel)
Kenji Sagae (University of Tokyo, Japan)
Sabine Schulte im Walde (University of Stuttgart, Germany)
Mark Steedman (University of Edinburgh, UK)
Suzanne Stevenson (University of Toronto, Canada)
Bert Vaux (University of Wisconsin, USA)
Charles Yang (University of Pennsylvania, USA)
Menno van Zaanen (Macquarie University, Australia)

Table of Contents

A Linguistic Investigation into Unsupervised DOP
Rens Bod 1

Using Classifier Features for Studying the Effect of Native Language on the Choice of Written Second Language Words
Oren Tsur and Ari Rappoport 9

Phon 1.2: A Computational Basis for Phonological Database Elaboration and Model Testing
Yvan Rose, Gregory Hedlund, Rod Byrne, Todd Wareham and Brian MacWhinney 17

High-accuracy Annotation and Parsing of CHILDES Transcripts
Kenji Sagae, Eric Davis, Alon Lavie, Brian MacWhinney and Shuly Wintner 25

I will shoot your shopping down and you can shoot all my tins—Automatic Lexical Acquisition from the CHILDES Database
Paula Buttery and Anna Korhonen 33

A Cognitive Model for the Representation and Acquisition of Verb Selectional Preferences
Afra Alishahi and Suzanne Stevenson 41

ISA meets Lara: An incremental word space model for cognitively plausible simulations of semantic learning
Marco Baroni, Alessandro Lenci and Luca Onnis 49

Simulating the acquisition of object names
Alessio Plebe, Vivian De La Cruz and Marco Mazzone 57

Rethinking the syntactic burst in young children
Christophe Parisse 65

The Topology of Synonymy and Homonymy Networks
James Gorman and James Curran 73

The Benefits of Errors: Learning an OT Grammar with a Structured Candidate Set
Tamas Biro 81

Learning to interpret novel noun-noun compounds: evidence from a category learning experiment
Barry Devereux and Fintan Costello 89

Conference Program

Friday, June 29, 2007

8:55–9:00	Opening Remarks
9:00–9:45	Invited Talk by Suzanne Stevenson
9.45–10:15	*A Linguistic Investigation into Unsupervised DOP* Rens Bod
10:15–10:45	*Using Classifier Features for Studying the Effect of Native Language on the Choice of Written Second Language Words* Oren Tsur and Ari Rappoport
	Morning Coffee Break
11.15–11.45	*Phon 1.2: A Computational Basis for Phonological Database Elaboration and Model Testing* Yvan Rose, Gregory Hedlund, Rod Byrne, Todd Wareham and Brian MacWhinney
11.45–12.15	*High-accuracy Annotation and Parsing of CHILDES Transcripts* Kenji Sagae, Eric Davis, Alon Lavie, Brian MacWhinney and Shuly Wintner
12.15–12.45	*I will shoot your shopping down and you can shoot all my tins—Automatic Lexical Acquisition from the CHILDES Database* Paula Buttery and Anna Korhonen
	Lunch
14.15–14.45	*A Cognitive Model for the Representation and Acquisition of Verb Selectional Preferences* Afra Alishahi and Suzanne Stevenson
14.45–15.15	*ISA meets Lara: An incremental word space model for cognitively plausible simulations of semantic learning* Marco Baroni, Alessandro Lenci and Luca Onnis
15.15–15.45	*Simulating the acquisition of object names* Alessio Plebe, Vivian De La Cruz and Marco Mazzone
	Afternoon Break

Friday, June 29, 2007 (continued)

16:15–16:45 *Rethinking the syntactic burst in young children*
Christophe Parisse

16:45–17:15 *The Topology of Synonymy and Homonymy Networks*
James Gorman and James Curran

17:15–17:45 *The Benefits of Errors: Learning an OT Grammar with a Structured Candidate Set*
Tamas Biro

17:45–18:15 *Learning to interpret novel noun-noun compounds: evidence from a category learning experiment*
Barry Devereux and Fintan Costello

18:15–18.20 Closing Remarks