

EACL 2012

**Workshop on Computational Models  
of Language Acquisition and Loss**

**Proceedings of the Conference**

April 24 2012  
Avignon France

© 2012 The Association for Computational Linguistics

ISBN 978-1-937284-19-0

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](mailto:acl@aclweb.org)

## Introduction

The past decades have seen a massive expansion in the application of statistical and machine learning methods to speech and natural language processing. This work has yielded impressive results which 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 loss.

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. On the other hand, changes in language abilities during aging and eventual losses related to conditions such as Alzheimer's disease and dementia have also attracted considerable investigative efforts. Parallels between the acquisition and loss have been raised, and a better understanding of the mechanisms involved in both, and of how the algorithms used to access concepts are affected in pathological cases can lead to earlier diagnosis and more targeted treatments.

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.

This was the third edition of this workshop that was first held at ACL 2007 in Prague and then in EACL 2009 in Athens. The workshop was 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. We invited submissions on relevant topics, including:

- Computational learning theory and analysis of language learning
- Computational models of first, second and bilingual language acquisition or of the evolution of language
- Computational models and analysis of factors that influence language acquisition and loss in different age groups and cultures
- Data resources and tools for investigating computational models of human language processes
- Empirical and theoretical comparisons of the environment and its impact on acquisition
- Investigations and comparisons of supervised, unsupervised and weakly-supervised methods for learning

Submissions included works on specific languages like English, Portuguese and Hebrew, and also crosslinguistic studies. Besides paper presentations the technical program included resources and systems demonstrations, and two invited talks by Mark Steedman, from University of Edinburgh (UK) and Alessandro Lenci, from University of Pisa (Italy).

## 1 Acknowledgments

We would like to thank the support of CNPq and CAPES-COFECUB (Projects 551964/2011-1, 478222/2011-4, 202007/2010-3 and 707/11) and of the labex (laboratoire d'excellence) Empirical Foundation of Linguistics.



**Organizers:**

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

**Program Committee:**

Afra Alishahi, Tilburg University (Netherlands)  
Colin J Bannard, University of Texas at Austin (USA)  
Marco Baroni, University of Trento (Italy)  
Jim Blevins, University of Cambridge (UK)  
Rens Bod, University of Amsterdam (Netherlands)  
Antal van den Bosch, Tilburg University (Netherlands)  
Alexander Clark, Royal Holloway, University of London (UK)  
Robin Clark, University of Pennsylvania (USA)  
Matthew W. Crocker, Saarland University (Germany)  
James Cussens, University of York (UK)  
Walter Daelemans, University of Antwerp (Belgium) and Tilburg University (Netherlands)  
Barry Devereux, University of Cambridge (UK)  
Sonja Eisenbeiss, University of Essex (UK)  
Afsaneh Fazly, University of Toronto (Canada)  
Cynthia Fisher, University of Illinois (USA)  
Jeroen Geertzen, University of Cambridge (UK)  
Henriette Hendriks, University of Cambridge (UK)  
Marco Idiart, Federal University of Rio Grande do Sul (Brazil)  
Aravind Joshi, University of Pennsylvania (USA)  
Shalom Lappin, King's College London (UK)  
Alessandro Lenci, University of Pisa (Italy)  
Igor Malioutov, Massachusetts Institute of Technology (USA)  
Marie-Catherine de Marneffe, Stanford University (USA)  
Fanny Meunier, Lumière Lyon 2 University (France)  
Brian Murphy, Carnegie Mellon University (USA)  
Maria Alice Parente, Federal University of Rio Grande do Sul (Brazil)  
Massimo Poesio, University of Essex (UK)  
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)  
Ekaterina Shutova, University of Cambridge (UK)  
Maity Siqueira, Federal University of Rio Grande do Sul (Brazil)  
Mark Steedman, University of Edinburgh (UK)

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)  
Michael Zock, LIF, CNRS, Marseille (France)

**Invited Speakers:**

Mark Steedman, University of Edinburgh (UK) and  
Alessandro Lenci, University of Pisa (Italy)

## Table of Contents

<i>Distinguishing Contact-Induced Change from Language Drift in Genetically Related Languages</i> T. Mark Ellison and Luisa Miceli .....	1
<i>Empiricist Solutions to Nativist Puzzles by means of Unsupervised TSG</i> Rens Bod and Margaux Smets .....	10
<i>Probabilistic Models of Grammar Acquisition</i> Mark Steedman .....	19
<i>A Morphologically Annotated Hebrew CHILDES Corpus</i> Aviad Albert, Brian MacWhinney, Bracha Nir and Shuly Wintner .....	20
<i>An annotated English child language database</i> Aline Villavicencio, Beracah Yankama, Rodrigo Wilkens, Marco Idiart and Robert Berwick .....	23
<i>Searching the Annotated Portuguese Childes Corpora</i> Rodrigo Wilkens .....	26
<i>Webservices for Bayesian Learning</i> Muntsa Padró and Núria Bel .....	29
<i>Unseen features. Collecting semantic data from congenital blind subjects</i> Alessandro Lenci, Marco Baroni and Giovanna Marotta .....	32
<i>PHACTS about activation-based word similarity effects</i> Basilio Calderone and Chiara Celata .....	33
<i>I say have you say tem: profiling verbs in children data in English and Portuguese</i> Rodrigo Wilkens and Aline Villavicencio .....	38
<i>Get out but don't fall down: verb-particle constructions in child language</i> Aline Villavicencio, Marco Idiart, Carlos Ramisch, Vítor Araújo, Beracah Yankama and Robert Berwick .....	43
<i>Phonologic Patterns of Brazilian Portuguese: a grapheme to phoneme converter based study</i> Vera Vasilévski .....	51





# Conference Program

**Tuesday, April 24, 2012**

8:55            Opening Session

**(9:00) Session 1: Language Evolution and Learning**

9:00            *Distinguishing Contact-Induced Change from Language Drift in Genetically Related Languages*

T. Mark Ellison and Luisa Miceli

9:30            *Empiricist Solutions to Nativist Puzzles by means of Unsupervised TSG*

Rens Bod and Margaux Smets

10:00          Coffee Break

**(10:30) Invited Talk I - Mark Steedman (University of Edinburgh, UK)**

*Probabilistic Models of Grammar Acquisition*

Mark Steedman

**(11:30) Session 2: Demonstrations**

*A Morphologically Annotated Hebrew CHILDES Corpus*

Aviad Albert, Brian MacWhinney, Bracha Nir and Shuly Wintner

*An annotated English child language database*

Aline Villavicencio, Beracah Yankama, Rodrigo Wilkens, Marco Idiart and Robert Berwick

*Searching the Annotated Portuguese Childes Corpora*

Rodrigo Wilkens

*Webservices for Bayesian Learning*

Muntsa Padró and Núria Bel

12:30          Lunch Break

**Tuesday, April 24, 2012 (continued)**

**(14:00) Invited Talk II - Alessandro Lenci (University of Pisa, Italy)**

*Unseen features. Collecting semantic data from congenital blind subjects*  
Alessandro Lenci, Marco Baroni and Giovanna Marotta

**(15:00) Session 3: Phonology and Syntax I**

15:00 *PHACTS about activation-based word similarity effects*  
Basilio Calderone and Chiara Celata

15:15 *I say have you say tem: profiling verbs in children data in English and Portuguese*  
Rodrigo Wilkens and Aline Villavicencio

15:30 Coffee Break

**(16:00) Session 4: Phonology and Syntax II**

16:00 *Get out but don't fall down: verb-particle constructions in child language*  
Aline Villavicencio, Marco Idiart, Carlos Ramisch, Vítor Araújo, Beracah Yankama and Robert Berwick

16:30 *Phonologic Patterns of Brazilian Portuguese: a grapheme to phoneme converter based study*  
Vera Vasilévski

17:00 Closing Remarks