

CMCL 2013

**Cognitive Modeling and Computational Linguistics**

**Proceedings of the Workshop**

August 8, 2013  
Sofia, Bulgaria

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

©2013 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](mailto:acl@aclweb.org)

ISBN 978-1-937284-61-9

## Introduction

The papers in these proceedings were presented at the Fourth Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL), held in Sofia, Bulgaria on 8 August 2013, in conjunction with the Annual Meeting of the Association for Computational Linguistics (ACL). The CMCL workshop series provides a unique venue for work on the interdisciplinary field of computational psycholinguistics, described by ACL Lifetime Achievement Award recipient Martin Kay as “build[ing] models of language that reflect in some interesting way on the ways in which people use language”. This workshop series builds on the tradition of earlier meetings, including the 1997 computational psycholinguistics meeting at the 1997 Annual Conference of the Cognitive Science Society in Berkeley, CA, and on the Incremental Parsing workshop held in 2004 at the Annual Meeting of the Association for Computational Linguistics.

We received nineteen submissions to the 2013 CMCL workshop, of which we accepted eleven for final appearance in the conference program. The overall quality of workshop submissions was extremely strong, reflecting the perennially increasing quality of work in this field. This year we also expanded the workshop by including keynote talks from two invited speakers, Dr Sharon Goldwater from the University of Edinburgh and Professor Rick Lewis from the University of Michigan, leading researchers in computational psycholinguistics. We would like to thank all submitting authors for allowing us to consider their work, the program committee for an outstanding job in reviewing and discussing submissions, and of course our invited speakers. We also gratefully acknowledge funding from the Cognitive Science Society for the Best Student Paper award, and to the Cluster of Excellence on “Multimodal Computing and Interaction” for assisting with funding of our invited speakers. Many thanks to all of you for your continued support of this workshop.

Vera Demberg and Roger Levy



**Organizers:**

Vera Demberg, Saarland University  
Roger Levy, UC San Diego

**Program Committee:**

Afra Alishahi, Tilburg University  
Klinton Bicknell, UC San Diego  
Matthew Crocker, Saarland University  
Brian Dillon, University of Massachusetts  
Afsaneh Fazly, University of Toronto  
Naomi Feldman, University of Maryland  
Michael C. Frank, Stanford University  
Stefan Frank, Radboud University Nijmegen  
Sharon Goldwater, Edinburgh University  
Noah Goodman, Stanford University  
John T. Hale, Cornell University  
T. Florian Jaeger, University of Rochester  
Frank Keller, University of Edinburgh  
Jeffrey Heinz, University of Delaware  
Richard L. Lewis, University of Michigan  
Brian Edmond Murphy, Carnegie Mellon University  
Timothy John O'Donnell, Massachusetts Institute of Technology  
Sebastian Padó, University of Heidelberg  
Ulrike Padó, Hochschule für Technik, Stuttgart  
Steven Piantadosi, University of Rochester  
David Reitter, Penn State University  
William Schuler, The Ohio State University  
Nathaniel Smith, University of Edinburgh  
Ed Stabler, UC Los Angeles  
Whitney Tabor, University of Connecticut and Haskins Laboratories

**Invited Speakers:**

Sharon Goldwater, University of Edinburgh  
Rick Lewis, University of Michigan



## Table of Contents

<i>Why is English so easy to segment?</i>	
Abdellah Fourtassi, Benjamin Börschinger, Mark Johnson and Emmanuel Dupoux . . . . .	1
<i>A model of generalization in distributional learning of phonetic categories</i>	
Bozena Pajak, Klinton Bicknell and Roger Levy . . . . .	11
<i>Learning non-concatenative morphology</i>	
Michelle Fullwood and Tim O'Donnell . . . . .	21
<i>Statistical Representation of Grammaticality Judgements: the Limits of N-Gram Models</i>	
Alexander Clark, Gianluca Giorgolo and Shalom Lappin . . . . .	28
<i>An Analysis of Memory-based Processing Costs using Incremental Deep Syntactic Dependency Parsing</i>	
Marten van Schijndel, Luan Nguyen and William Schuler . . . . .	37
<i>Computational simulations of second language construction learning</i>	
Yevgen Matuselych, Afra Alishahi and Ad Backus . . . . .	47
<i>The semantic augmentation of a psycholinguistically-motivated syntactic formalism</i>	
Asad Sayeed and Vera Demberg . . . . .	57
<i>Evaluating Neighbor Rank and Distance Measures as Predictors of Semantic Priming</i>	
Gabriella Lapesa and Stefan Evert . . . . .	66
<i>Concreteness and Corpora: A Theoretical and Practical Study</i>	
Felix Hill, Douwe Kiela and Anna Korhonen . . . . .	75
<i>On the Information Conveyed by Discourse Markers</i>	
Fatemeh Torabi Asr and Vera Demberg . . . . .	84
<i>Incremental Grammar Induction from Child-Directed Dialogue Utterances</i>	
Arash Eshghi, Julian Hough and Matthew Purver . . . . .	94





# Workshop Program

**Thursday, August 8th, 2013**

8:25            Opening Remarks

8:30            Invited Talk by Sharon Goldwater

## **Session 1: Segmentation and Phonetics**

09:30            *Why is English so easy to segment?*  
Abdellah Fourtassi, Benjamin Börschinger, Mark Johnson and Emmanuel Dupoux

10:00            *A model of generalization in distributional learning of phonetic categories*  
Bozena Pajak, Klinton Bicknell and Roger Levy

10:30            Coffee break

## **Session 2: Syntax and Morphology**

11:00            *Learning non-concatenative morphology*  
Michelle Fullwood and Tim O'Donnell

11:30            *Statistical Representation of Grammaticality Judgements: the Limits of N-Gram Models*  
Alexander Clark, Gianluca Giorgolo and Shalom Lappin

12:00            *An Analysis of Memory-based Processing Costs using Incremental Deep Syntactic Dependency Parsing*  
Marten van Schijndel, Luan Nguyen and William Schuler

12:30            *Computational simulations of second language construction learning*  
Yevgen Matuskevych, Afra Alishahi and Ad Backus

13:00            Lunch break

**Thursday, August 8th, 2013 (continued)**

**Session 3: Semantics**

14:00 *The semantic augmentation of a psycholinguistically-motivated syntactic formalism*  
Asad Sayeed and Vera Demberg

14:30 *Evaluating Neighbor Rank and Distance Measures as Predictors of Semantic Priming*  
Gabriella Lapesa and Stefan Evert

15:00 *Concreteness and Corpora: A Theoretical and Practical Study*  
Felix Hill, Douwe Kiela and Anna Korhonen

15:30 Coffee break

**Session 4: Discourse and Dialog**

16:00 *On the Information Conveyed by Discourse Markers*  
Fatemeh Torabi Asr and Vera Demberg

16:30 *Incremental Grammar Induction from Child-Directed Dialogue Utterances*  
Arash Eshghi, Julian Hough and Matthew Purver

17:00 Invited Talk by Rick Lewis