

ACL 2014

**Workshop on Language Technologies  
and Computational Social Science**

**Proceedings of the Workshop**

June 26, 2014  
Baltimore, Maryland, USA



©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](mailto:acl@aclweb.org)

ISBN 978-1-941643-10-5

## Preface

The ACL 2014 Workshop on Language Technologies and Computational Social Science was held on June 26, 2014 in Baltimore, following the 52nd annual meeting of the ACL. The workshop's goal was to increase the visibility of computational social science— in which automated techniques are applied to massive datasets to answer scientific questions about society—for ACL researchers and to help build connections between language technologists and social scientists.

The workshop included six invited talks from researchers who have successfully brought language technologies to computational social science research questions: political scientists Amber Boydston and Justin Grimmer, social computing expert Ed Chi, sociolinguist Sali Tagliamonte, and computational linguists Lillian Lee and Philip Resnik.

Out of twenty submissions of short papers, thirteen were selected by the program committee for poster presentation. These represent an exciting, conversation-provoking range of research projects.

The workshop also included a session reporting on a related research competition, the NLP Unshared Task in PoliInformatics. A short overview paper is included in these proceedings.

This workshop was supported by grants from Google and the U.S. National Science Foundation (grant IIS-1433108). These funds enabled the participation of the invited speakers, fourteen graduate students, and three more senior researchers.

We thank the invited speakers, program committee members, authors, and participants for sharing time and thoughts on this increasingly important research topic.

Cristian Danescu-Niculescu-Mizil, Jacob Eisenstein, Kathy McKeown, and Noah Smith



**Organizers:**

Cristian Danescu-Niculescu-Mizil, Max Planck Institute SWS  
Jacob Eisenstein, Georgia Institute of Technology  
Kathleen McKeown, Columbia University  
Noah A. Smith, Carnegie Mellon University

**Program Committee:**

Jordan Boyd-Graber, University of Maryland  
Claire Cardie, Cornell University  
Munmun de Choudhury, Georgia Institute of Technology  
Cindy Chung, University of Texas  
Mark Dredze, Johns Hopkins University  
Dan Jurafsky, Stanford University  
Maria Liakata, University of Warwick  
Brendan O'Connor, Carnegie Mellon University  
Bo Pang, Google  
Daniel Preoțiuc-Pietro, Sheffield University  
Owen Rambow, Columbia University  
Jaime Teevan, Microsoft  
Hanna Wallach, University of Massachusetts

**Invited Speakers:**

Amber Boydstun, University of California at Davis  
Ed Chi, Google  
Justin Grimmer, Stanford University  
Lillian Lee, Cornell University  
Philip Resnik, University of Maryland  
Sali Tagliamonte, University of Toronto



## Table of Contents

<i>Is It All in the Phrasing? Computational Explorations in How We Say What We Say, and Why It Matters</i> Lillian Lee .....	1
<i>Creating and Destroying Party Brands</i> Justin Grimmer .....	2
<i>Sociolinguistics for Computational Social Science</i> Sali Tagliamonte .....	3
<i>Location and Language Use in Social Media</i> Ed Chi .....	4
<i>Overview of the 2014 NLP Unshared Task in PoliInformatics</i> Noah A. Smith, Claire Cardie, Anne Washington and John Wilkerson .....	5
<i>Context-based Natural Language Processing for GIS-based Vague Region Visualization</i> Wei Chen .....	8
<i>Extracting Socioeconomic Patterns from the News: Modelling Text and Outlet Importance Jointly</i> Vasileios Lampos, Daniel Preoțiu-Pietro, Sina Samangooei, Douwe Gelling and Trevor Cohn .	13
<i>Fact Checking: Task definition and dataset construction</i> Andreas Vlachos and Sebastian Riedel .....	18
<i>Finding Eyewitness Tweets During Crises</i> Fred Morstatter, Nichola Lubold, Heather Pon-Barry, Jürgen Pfeffer and Huan Liu .....	23
<i>Inducing Information Structures for Data-driven Text Analysis</i> Andrew Salway, Samia Touileb and Endre Tvinnereim .....	28
<i>Information density, Heaps' Law, and perception of factiness in news</i> Miriam Boon .....	33
<i>Measuring the Public Accountability of New Modes of Governance</i> Bruno Wueest, Gerold Schneider and Michael Amsler .....	38
<i>Optimizing Features in Active Machine Learning for Complex Qualitative Content Analysis</i> Jasy Suet Yan Liew, Nancy McCracken, Shichun Zhou and Kevin Crowston .....	44
<i>Power of Confidence: How Poll Scores Impact Topic Dynamics in Political Debates</i> Vinodkumar Prabhakaran, Ashima Arora and Owen Rambow .....	49
<i>Predicting Fine-grained Social Roles with Selectional Preferences</i> Charley Beller, Craig Harman and Benjamin Van Durme .....	50
<i>Predicting Party Affiliations from European Parliament Debates</i> Bjørn Høyland, Jean-François Godbout, Emanuele Laponi and Erik Velldal .....	56
<i>Temporal Analysis of Language through Neural Language Models</i> Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde and Slav Petrov .....	61
<i>Using Simple NLP Tools to Trace the Globalization of the Art World</i> Mohamed AlTantawy, Alix Rule, Owen Rambow, Zhongyu Wang and Rupayan Basu .....	66

*Issue Framing as a Generalizable Phenomenon*  
Amber Boydston ..... 71

*“I Want to Talk About, Again, My Record On Energy ...”: Modeling Agendas and Framing in Political Debates and Other Conversations*  
Philip Resnik..... 72



# Workshop Program

**Thursday, June 26, 2014**

## **Welcome; Invited Talks, Style and Rhetoric**

9:00 Welcome

9:05 *Is It All in the Phrasing? Computational Explorations in How We Say What We Say, and Why It Matters*  
Lillian Lee

9:35 *Creating and Destroying Party Brands*  
Justin Grimmer

10:05 Discussion

**(10:30-11:00) Coffee break**

## **Invited Talks, Sociolinguistics and Social Media**

11:00 *Sociolinguistics for Computational Social Science*  
Sali Tagliamonte

11:30 *Location and Language Use in Social Media*  
Ed Chi

12:00 Discussion

**Thursday, June 26, 2014 (continued)**

**(12:30-2:00) Lunch break**

**(2:00-2:30) Unshared task**

2:00

*Overview of the 2014 NLP Unshared Task in PoliInformatics*

Noah A. Smith, Claire Cardie, Anne Washington and John Wilkerson

**(2:30-3:30) Poster session**

*Context-based Natural Language Processing for GIS-based Vague Region Visualization*

Wei Chen

*Extracting Socioeconomic Patterns from the News: Modelling Text and Outlet Importance Jointly*

Vasileios Lampos, Daniel Preoȃiuc-Pietro, Sina Samangooei, Douwe Gelling and Trevor Cohn

*Fact Checking: Task definition and dataset construction*

Andreas Vlachos and Sebastian Riedel

*Finding Eyewitness Tweets During Crises*

Fred Morstatter, Nichola Lubold, Heather Pon-Barry, Jürgen Pfeffer and Huan Liu

*Inducing Information Structures for Data-driven Text Analysis*

Andrew Salway, Samia Touileb and Endre Tvinnereim

*Information density, Heaps' Law, and perception of factiness in news*

Miriam Boon

*Measuring the Public Accountability of New Modes of Governance*

Bruno Wueest, Gerold Schneider and Michael Amsler

*Optimizing Features in Active Machine Learning for Complex Qualitative Content Analysis*

Jasy Suet Yan Liew, Nancy McCracken, Shichun Zhou and Kevin Crowston

*Power of Confidence: How Poll Scores Impact Topic Dynamics in Political Debates*

Vinodkumar Prabhakaran, Ashima Arora and Owen Rambow

**Thursday, June 26, 2014 (continued)**

*Predicting Fine-grained Social Roles with Selectional Preferences*

Charley Beller, Craig Harman and Benjamin Van Durme

*Predicting Party Affiliations from European Parliament Debates*

Bjørn Høyland, Jean-François Godbout, Emanuele Lapponi and Erik Velldal

*Temporal Analysis of Language through Neural Language Models*

Yoon Kim, Yi-I Chiu, Kentaro Hanaki, Darshan Hegde and Slav Petrov

*Using Simple NLP Tools to Trace the Globalization of the Art World*

Mohamed AlTantawy, Alix Rule, Owen Rambow, Zhongyu Wang and Rupayan Basu

**(3:30-4:00) Coffee break**

**Invited Talks, Framing and Agenda-Setting**

4:00

*Issue Framing as a Generalizable Phenomenon*

Amber Boydston

4:30

*“I Want to Talk About, Again, My Record On Energy ...”: Modeling Agendas and Framing in Political Debates and Other Conversations*

Philip Resnik

5:00

Discussion

**(5:30) Closing remarks**

