

ACL 2017

**The 55th Annual Meeting of the  
Association for Computational Linguistics**

**Proceedings of the Second Workshop on Natural Language  
Processing and Computational Social Science**

August 3, 2017  
Vancouver, Canada

©2017 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 978-1-945626-65-4

## Introduction

Welcome to the Second Workshop on NLP and Computational Social Science!

After a successful installation last year, we again received a large number of high-quality submission this year, an indication that interest in the topic is growing. We received 31 submissions, and due to a rigorous review process by our committee, we accepted 18. The program this year includes 4 papers presented as spotlight talks, and 14 posters.

We are especially excited to see so many submissions from outside of NLP, and hope to continue the tradition to foster a dialogue between NLP researchers and users of NLP technology in the social sciences.

We are also glad to present a fantastic selection of invited speakers from various aspects of computational social science.

We would like to thank all authors of the accepted papers, our invited speakers, and the fantastic organizing committee that made this workshop possible, and, last but not least, all attendees!

The NLP and CSS workshop organizing team



**Organizers:**

Dirk Hovy (University of Copenhagen), Svitlana Volkova (PNNL), David Bamman (UC Berkeley), David Jurgens (Stanford), Brendan O'Connor (UMass Amherst), Oren Tsur (Harvard/Northeastern) and A. Seza Doğruöz (independent researcher)

**Program Committee:**

Natalie Ahn  
Nikolaos Aletras  
Kristen M. Altenburger  
Tim Althoff  
Timothy Baldwin  
John Beierl  
Eric Bell  
Reihane Boghrati  
Kalina Bontcheva  
Ryan Boyd  
Kaiping Chen  
Munmun De Choudhury  
Matthew Denny  
Valery Dzutsati  
Jacob Eisenstein  
Jesse Freitas  
Justin Garten  
Dan Goldwasser  
William L. Hamilton  
Oul Han  
Abram Handler  
Marti A. Hearst  
Graeme Hirst  
Eduard Hovy  
Loring Ingraham  
Kristen Johnson  
Kenneth Joseph  
Sunghwan Mac Kim  
Vivek Kulkarni  
Vasileios Lamos  
Corey Lester  
Kate Loveys  
Peter Makarov  
Rada Mihalcea  
Slava Mikhaylov  
Dong Nguyen  
Alice Oh  
Cecile Paris  
Michael J. Paul  
Thierry Poibeau  
Christopher Potts

Adam Poulston  
Vinodkumar Prabhakaran  
Daniel Preoŕiuc-Pietro  
Sravana Reddy  
Rebecca Resnik  
Ludovic Rheault  
Molly Roberts  
Pedro Rodriguez  
Carolyn Rose  
Sara Rosenthal  
Derek Ruths  
Fiona Shen-Bayh  
Mark Stevenson  
Brandon Stewart  
Ian Stewart  
Anders Sŕgaard  
Rachael Tatman  
Rob Voigt  
Stephen Wan  
Alex Wang  
ZeeraK Waseem  
Swede White  
Steven Wilson

**Invited Speaker:**

Lyle Ungar, University of Pennsylvania  
Gideon Mann, Bloomberg  
Brandon Stewart, Princeton

## Table of Contents

|  |    |
|--|----|
| <i>Language-independent Gender Prediction on Twitter</i><br>Nikola Ljubešić, Darja Fišer and Tomaž Erjavec .....   | 1  |
| <i>When does a compliment become sexist? Analysis and classification of ambivalent sexism using twitter data</i><br>Akshita Jha and Radhika Mamidi .....                                   | 7  |
| <i>Personality Driven Differences in Paraphrase Preference</i><br>Daniel Preoțiuc-Pietro, Jordan Carpenter and Lyle Ungar .....  | 17 |
| <i>community2vec: Vector representations of online communities encode semantic relationships</i><br>Trevor Martin .....  | 27 |
| <i>Telling Apart Tweets Associated with Controversial versus Non-Controversial Topics</i><br>Aseel Addawood, Rezvaneh Rezapour, Omid Abdar and Jana Diesner .....                          | 32 |
| <i>Cross-Lingual Classification of Topics in Political Texts</i><br>Goran Glavaš, Federico Nanni and Simone Paolo Ponzetto .....   | 42 |
| <i>Mining Social Science Publications for Survey Variables</i><br>Andrea Zielinski and Peter Mutschke .....  | 47 |
| <i>Linguistic Markers of Influence in Informal Interactions</i><br>Shrimai Prabhumoye, Samridhi Choudhary, Evangelia Spiliopoulou, Christopher Bogart, Carolyn Rose and Alan W Black ..... | 53 |
| <i>Non-lexical Features Encode Political Affiliation on Twitter</i><br>Rachael Tatman, Leo Stewart, Amandalynne Paullada and Emma Spiro .....  | 63 |
| <i>Modelling Participation in Small Group Social Sequences with Markov Rewards Analysis</i><br>Gabriel Murray .....  | 68 |
| <i>Code-Switching as a Social Act: The Case of Arabic Wikipedia Talk Pages</i><br>Michael Yoder, Shruti Rijhwani, Carolyn Rosé and Lori Levin .....  | 73 |
| <i>How Does Twitter User Behavior Vary Across Demographic Groups?</i><br>Zach Wood-Doughty, Michael Smith, David Broniatowski and Mark Dredze .....  | 83 |
| <i>Ideological Phrase Indicators for Classification of Political Discourse Framing on Twitter</i><br>Kristen Johnson, I-Ta Lee and Dan Goldwasser .....                                    | 90 |



# Workshop Program

**Aug-3-2017**

**9:00–10:30**    **Session 1**

9:00–9:15    *Welcome*  
Organizers

**9:15–10:00**    *Invited Talk 1*

**10:00–10:30**    **Spotlight Paper Session**

10:00–10:15    *Language-independent Gender Prediction on Twitter*  
Nikola Ljubešić, Darja Fišer and Tomaž Erjavec

10:15–10:30    *When does a compliment become sexist? Analysis and classification of ambivalent sexism using twitter data*  
Akshita Jha and Radhika Mamidi

**10:30–11:00**    *Morning coffee break*

**11:00–12:15**    **Session 2**

11:00–11:45    *Invited Talk: Measuring Psychological Traits using Social Media*  
Lyle Ungar

**Aug-3-2017 (continued)**

**11:45–12:15 Spotlight Paper Session**

11:45–12:00 *Personality Driven Differences in Paraphrase Preference*  
Daniel Preoțiu-Pietro, Jordan Carpenter and Lyle Ungar

12:00–12:15 *Never Tell Me the Odds: How Belief Dynamics Shape Audience Experience (non-archival)*  
Shengli Hu

**12:15–14:00 Lunch break**

**14:00–15:30 Session 3**

14:00–14:45 *Invited Talk: The War on Facts*  
Gideon Mann

**14:45–15:30 One-minute-madness paper presentation**

14:45–14:48 *community2vec: Vector representations of online communities encode semantic relationships*  
Trevor Martin

14:48–14:51 *Telling Apart Tweets Associated with Controversial versus Non-Controversial Topics*  
Aseel Addawood, Rezvaneh Rezapour, Omid Abdar and Jana Diesner

14:51–14:54 *Cross-Lingual Classification of Topics in Political Texts*  
Goran Glavaš, Federico Nanni and Simone Paolo Ponzetto

14:54–14:57 *The Role of Network Structure for Gender Prediction (non-archival)*  
Kristen M. Altenburger and Johan Ugander

14:57–15:00 *The Role of Party and Incumbency in Identification of Argumentation Strategies in Political Debate (non-archival)*  
Justin Garten, Kenji Sagae, Zahra Kamel, Nitika Awasthi and Morteza Dehghani

15:00–15:03 *Mining Social Science Publications for Survey Variables*  
Andrea Zielinski and Peter Mutschke

**Aug-3-2017 (continued)**

- 15:03–15:06 *Linguistic Markers of Influence in Informal Interactions*  
Shrimai Prabhume, Samridhi Choudhary, Evangelia Spiliopoulou, Christopher Bogart, Carolyn Rose and Alan W Black
- 15:06–15:09 *Non-lexical Features Encode Political Affiliation on Twitter*  
Rachael Tatman, Leo Stewart, Amandalynne Paullada and Emma Spiro
- 15:09–15:12 *Syntactic Alignment in Power Relations (non-archival)*  
Reihane Boghrati, Justin Garten and Morteza Dehghani
- 15:12–15:15 *Modelling Participation in Small Group Social Sequences with Markov Rewards Analysis*  
Gabriel Murray
- 15:15–15:18 *Code-Switching as a Social Act: The Case of Arabic Wikipedia Talk Pages*  
Michael Yoder, Shruti Rijhwani, Carolyn Rosé and Lori Levin
- 15:18–15:21 *How Does Twitter User Behavior Vary Across Demographic Groups?*  
Zach Wood-Doughty, Michael Smith, David Broniatowski and Mark Dredze
- 15:21–15:24 *Ideological Phrase Indicators for Classification of Political Discourse Framing on Twitter*  
Kristen Johnson, I-Ta Lee and Dan Goldwasser
- 15:24–15:27 *Market Evolution of Sharing Economy vs. Traditional Platforms: A Natural Language Processing Approach*  
Mohsen Mosleh and Babak Heydari
- 15:30–16:45** *Coffee break and posters*

**Aug-3-2017 (continued)**

**16:45–17:45 Session 4**

16:45–17:30 *Invited Talk*  
Brandon Stewart

17:30–17:45 *Closing remarks and wrap-up*  
Organizers