

NLP+CSS 2020

**Natural Language Processing  
and Computational Social Sciences**

**Proceedings of the Fourth Workshop**

November 20, 2020

*Online*

©2020 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 978-1-952148-80-4

## Introduction

Welcome to the Fourth Workshop on Natural Language Processing (NLP) and Computational Social Science (CSS)! This workshop series builds on a successful string of iterations, with dozens of interdisciplinary submissions to make NLP techniques and insights standard practice in CSS research—as well as improve NLP through insights from Social Science.

We received a record 61 submissions and after a rigorous review process by our committee, we accepted 23 archival entries, and 4 non-archival abstracts. In a time of uncertainty and a global pandemic, we are heartened to see strong interest and genuinely excited to share these fantastic works. Due to the conference shifting to a virtual format, all papers will be presented with pre-recorded videos during two sessions of the workshop, allowing authors to engage through chat and video throughout the workshop. We are especially excited to see so many submissions from outside of NLP, and hope to continue the tradition to foster a dialogue between researchers in NLP and these other fields.

We are glad to present a fantastic selection of invited speakers from various aspects of computational social science: Elizabeth E. Bruch (University of Michigan), Dong Nguyen (Utrecht University), Jesse Shapiro (Brown University), and Diyi Yang (Georgia Institute of Technology). This year, we recognize the growing role of NLP+CSS technology outside of academia and aimed to highlight opportunities for students and practitioners outside of academia. To help this effort, we are thrilled to host a panel discuss NLP and CSS in industry with amazing speakers from a variety of tech companies: Kristin Althenburger (Facebook), Jason Baldrige (Google), Glen Coppersmith (Qntfy), Emre Kiciman (Microsoft Research), and Umashanthi Pavalanathan (Twitter).

Recognizing the growing community of CSS researchers outside of NLP, as well as new researchers starting out in NLP+CSS, we introduced a new mentoring program that connected volunteers from our Program Committee to authors wanting additional guidance. We are especially grateful for their time and guidance.

We would like to thank the Program Committee members who reviewed the papers this year. We would also like to thank the workshop participants. Last, a word of thanks also goes to our sponsors Microsoft Research and Facebook, who enabled us to support a record number of students this year.

David Bamman, Dirk Hovy, David Jurgens, Brendan O'Connor, and Svitlana Volkova (Co-Organizers)



**Organizers:**

David Bamman, University of California, Berkeley  
Dirk Hovy, Bocconi University  
David Jurgens, University of Michigan  
Brendan O'Connor, University of Massachusetts Amherst  
Svitlana Volkova, Pacific Northwest National Laboratory

**Program Committee:**

Natalie Ahn  
Nikolaos Aletras  
Kristen M. Altenburger  
Vijjini Anvesh Rao  
Alexandra Balahur  
Timothy Baldwin  
Alon Bartal  
Eric Bell  
Laura Biester  
Michael Bloodgood  
Chris Brew  
Julian Brooke  
Aron Culotta  
Steve DeNeeffe  
Matthew Denny  
A. Seza Dođruöz  
Jonathan Dunn  
Valery Dzutsati  
Jacob Eisenstein  
Jason Eisner  
Micha Elsner  
Alex Fine  
Tommaso Fornaciari  
Kathleen C. Fraser  
Jesse Freitas  
Juri Ganitkevitch  
Dan Goldwasser  
Yvette Graham  
Andrew Halterman  
Oul Han  
Marti A. Hearst  
Michael Heilman  
Loring Ingraham  
Molly Ireland  
Anders Johannsen  
Kristen Johnson  
Kenneth Joseph  
Brendan Kennedy  
Sunghwan Mac Kim  
Vivek Kulkarni

Jonathan K. Kummerfeld  
Alessandro Lenci  
Mingyang Li  
Nikola Ljubešić  
Josephine Lukito  
Teresa Lynn  
Elijah Mayfield  
Diana Maynard  
Paul McLachlan  
Julia Mendelsohn  
Ben Miller  
David Mimno  
Shubhanshu Mishra  
Aida Mostafazadeh Davani  
Dong Nguyen  
Pierre Nugues  
Stephan Oepen  
Miles Osborne  
Sebastian Padó  
Shriphani Palakodety  
Ted Pedersen  
Viktor Pekar  
Barbara Plank  
Massimo Poesio  
Thierry Poibeau  
Daniel Preoțiuc-Pietro  
Afshin Rahimi  
Sravana Reddy  
Ellen Riloff  
Anthony Rios  
Carolyn Rosé  
Alla Rozovskaya  
Derek Ruths  
Asad Sayeed  
Natalie Schluter  
Djamé Seddah  
Felix Soldner  
Vivek Srikumar  
Ian Stewart  
Oren Tsur  
Lyle Ungar  
Esther van den Berg  
Zijian Wang  
Swede White  
Steven Wilson  
Wei Xu  
Michael Yeomans  
Shuang (Sophie) Zhai

**Invited Speakers:**

Elizabeth E. Bruch, University of Michigan  
Dong Nguyen, Utrecht University  
Jesse Shapiro, Brown University  
Diyi Yang, Georgia Institute of Technology

**Panelists:**

Kristin Althenburger, Facebook  
Jason Baldrige, Google  
Glen Coppersmith, Qntfy  
Emre Kiciman, Microsoft Research  
Umashanthi Pavalanathan, Twitter





## Table of Contents

<i>Measuring Linguistic Diversity During COVID-19</i> Jonathan Dunn, Tom Coupe and Benjamin Adams .....	1
<i>Using BERT for Qualitative Content Analysis in Psychosocial Online Counseling</i> Philipp Grandeit, Carolyn Haberkern, Maximiliane Lang, Jens Albrecht and Robert Lehmann ..	11
<i>Swimming with the Tide? Positional Claim Detection across Political Text Types</i> Nico Blokker, Erenay Dayanik, Gabriella Lapesa and Sebastian Padó .....	24
<i>Does Social Support (Expressed in Post Titles) Elicit Comments in Online Substance Use Recovery Forums?</i> Anietie Andy and Sharath Chandra Guntuku .....	35
<i>I miss you babe: Analyzing Emotion Dynamics During COVID-19 Pandemic</i> Hui Xian Lynnette Ng, Roy Ka-Wei Lee and Md Rabiul Awal .....	41
<i>Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data</i> Alex Fine, Patrick Crutchley, Jenny Blase, Joshua Carroll and Glen Coppersmith .....	50
<i>Viable Threat on News Reading: Generating Biased News Using Natural Language Models</i> Saurabh Gupta, Hong Huy Nguyen, Junichi Yamagishi and Isao Echizen .....	55
<i>Unsupervised Anomaly Detection in Parole Hearings using Language Models</i> Graham Todd, Catalin Voss and Jenny Hong .....	66
<i>Identifying Worry in Twitter: Beyond Emotion Analysis</i> Reyha Verma, Christian von der Weth, Jithin Vachery and Mohan Kankanhalli .....	72
<i>Text Zoning and Classification for Job Advertisements in German, French and English</i> Ann-Sophie Gnehm and Simon Clematide .....	83
<i>Is Wikipedia succeeding in reducing gender bias? Assessing changes in gender bias in Wikipedia using word embeddings</i> Katja Geertruida Schmahl, Tom Julian Viering, Stavros Makrodimitris, Arman Naseri Jahfari, David Tax and Marco Loog .....	94
<i>Effects of Anonymity on Comment Persuasiveness in Wikipedia Articles for Deletion Discussions</i> Yimin Xiao and Lu Xiao .....	104
<i>Uncertainty over Uncertainty: Investigating the Assumptions, Annotations, and Text Measurements of Economic Policy Uncertainty</i> Katherine Keith, Christoph Teichmann, Brendan O'Connor and Edgar Meij .....	116
<i>Recalibrating classifiers for interpretable abusive content detection</i> Bertie Vidgen, Scott Hale, Sam Staton, Tom Melham, Helen Margetts, Ohad Kammar and Marcin Szymczak .....	132
<i>Predicting independent living outcomes from written reports of social workers</i> Angelika Maier and Philipp Cimiano .....	139

<i>Analyzing Political Bias and Unfairness in News Articles at Different Levels of Granularity</i> Wei-Fan Chen, Khalid Al Khatib, Henning Wachsmuth and Benno Stein .....	149
<i>Mapping Local News Coverage: Precise location extraction in textual news content using fine-tuned BERT based language model</i> Sarang Gupta and Kumari Nishu .....	155
<i>Foreigner-directed speech is simpler than native-directed: Evidence from social media</i> Aleksandrs Berdicevskis .....	163
<i>Diachronic Embeddings for People in the News</i> Felix Hennig and Steven Wilson .....	173
<i>Social media data as a lens onto care-seeking behavior among women veterans of the US armed forces</i> Kacie Kelly, Alex Fine and Glen Coppersmith .....	184
<i>Understanding Weekly COVID-19 Concerns through Dynamic Content-Specific LDA Topic Modeling</i> Mohammadzaman Zamani, H. Andrew Schwartz, Johannes Eichstaedt, Sharath Chandra Guntuku, Adithya Virinchipuram Ganesan, Sean Clouston and Salvatore Giorgi .....	193
<i>Emoji and Self-Identity in Twitter Bios</i> Jinhang Li, Giorgos Longinos, Steven Wilson and Walid Magdy .....	199
<i>Analyzing Gender Bias within Narrative Tropes</i> Dhruvil Gala, Mohammad Omar Khursheed, Hannah Lerner, Brendan O'Connor and Mohit Iyyer	212

# Conference Program

Friday, November 20, 2020 (All times PST; UTC-8)

**07:00–07:15** *Introductory Remarks*

**07:15–09:00** **Session 1**

07:15–08:00 *Invited Talk 1: When NLP Meets Language Variation*  
Dong Nguyen

08:00–09:00 *Paper Session and Discussion 1*  
All papers

08:00–09:00 *Measuring Linguistic Diversity During COVID-19*  
Jonathan Dunn, Tom Coupe and Benjamin Adams

08:00–09:00 *How Language Influences Attitudes Toward Brands*  
David DeFranza, Arul Mishra and Himanshu Mishra

08:00–09:00 *Using BERT for Qualitative Content Analysis in Psychosocial Online Counseling*  
Philipp Grandeit, Carolyn Haberkern, Maximiliane Lang, Jens Albrecht and Robert Lehmann

08:00–09:00 *Swimming with the Tide? Positional Claim Detection across Political Text Types*  
Nico Blokker, Erenay Dayanik, Gabriella Lapesa and Sebastian Padó

08:00–09:00 *Does Social Support (Expressed in Post Titles) Elicit Comments in Online Substance Use Recovery Forums?*  
Anietie Andy and Sharath Chandra Guntuku

08:00–09:00 *I miss you babe: Analyzing Emotion Dynamics During COVID-19 Pandemic*  
Hui Xian Lynnette Ng, Roy Ka-Wei Lee and Md Rabiul Awal

08:00–09:00 *Assessing population-level symptoms of anxiety, depression, and suicide risk in real time using NLP applied to social media data*  
Alex Fine, Patrick Crutchley, Jenny Blase, Joshua Carroll and Glen Coppersmith

08:00–09:00 *Topic preference detection: A novel approach to understand perspective taking in conversation*  
Michael Yeomans and Alison Wood Brooks

**Friday, November 20, 2020 (All times PST; UTC-8) (continued)**

- 08:00–09:00 *Viable Threat on News Reading: Generating Biased News Using Natural Language Models*  
Saurabh Gupta, Hong Huy Nguyen, Junichi Yamagishi and Isao Echizen
- 08:00–09:00 *A Lexical Semantic Leadership Network of Nineteenth Century Abolitionist Newspapers*  
Sandeep Soni, Lauren Klein and Jacob Eisenstein
- 08:00–09:00 *Unsupervised Anomaly Detection in Parole Hearings using Language Models*  
Graham Todd, Catalin Voss and Jenny Hong
- 08:00–09:00 *Identifying Worry in Twitter: Beyond Emotion Analysis*  
Reyha Verma, Christian von der Weth, Jithin Vachery and Mohan Kankanhalli
- 8:00–9:00 *Text Zoning and Classification for Job Advertisements in German, French and English*  
Ann-Sophie Gnehm and Simon Clematide
- 08:00–09:00 *Is Wikipedia succeeding in reducing gender bias? Assessing changes in gender bias in Wikipedia using word embeddings*  
Katja Geertruida Schmahl, Tom Julian Viering, Stavros Makrodimitris, Arman Naseri Jahfari, David Tax and Marco Loog
- 08:00–09:00 *Effects of Anonymity on Comment Persuasiveness in Wikipedia Articles for Deletion Discussions*  
Yimin Xiao and Lu Xiao
- 08:00–09:00 *Uncertainty over Uncertainty: Investigating the Assumptions, Annotations, and Text Measurements of Economic Policy Uncertainty*  
Katherine Keith, Christoph Teichmann, Brendan O’Connor and Edgar Meij
- 08:00–09:00 *Recalibrating classifiers for interpretable abusive content detection*  
Bertie Vidgen, Scott Hale, Sam Staton, Tom Melham, Helen Margetts, Ohad Kammar and Marcin Szymczak
- 08:00–09:00 *Predicting independent living outcomes from written reports of social workers*  
Angelika Maier and Philipp Cimiano
- 08:00–09:00 *Analyzing Political Bias and Unfairness in News Articles at Different Levels of Granularity*  
Wei-Fan Chen, Khalid Al Khatib, Henning Wachsmuth and Benno Stein
- 08:00–09:00 *Mapping Local News Coverage: Precise location extraction in textual news content using fine-tuned BERT based language model*  
Sarang Gupta and Kumari Nishu

**Friday, November 20, 2020 (All times PST; UTC-8) (continued)**

- 08:00–09:00 *Foreigner-directed speech is simpler than native-directed: Evidence from social media*  
Aleksandrs Berdicevskis
- 08:00–09:00 *Diachronic Embeddings for People in the News*  
Felix Hennig and Steven Wilson
- 08:00–09:00 *Social media data as a lens onto care-seeking behavior among women veterans of the US armed forces*  
Kacie Kelly, Alex Fine and Glen Coppersmith
- 08:00–09:00 *Understanding Weekly COVID-19 Concerns through Dynamic Content-Specific LDA Topic Modeling*  
Mohammadzaman Zamani, H. Andrew Schwartz, Johannes Eichstaedt, Sharath Chandra Guntuku, Adithya Virinchipuram Ganesan, Sean Clouston and Salvatore Giorgi
- 08:00–09:00 *Emoji and Self-Identity in Twitter Bios*  
Jinhang Li, Giorgos Longinos, Steven Wilson and Walid Magdy
- 08:00–09:00 *Analyzing Gender Bias within Narrative Tropes*  
Dhruvil Gala, Mohammad Omar Khursheed, Hannah Lerner, Brendan O’Connor and Mohit Iyyer
- 08:00–09:00 *An Unfair Affinity Toward Fairness: Characterizing 70 Years of Social Biases in B<sup>H</sup>ollywood*  
Kunal Khadilkar and Ashiqur KhudaBukhsh
- 09:00–09:15 Break**

**Friday, November 20, 2020 (All times PST; UTC-8) (continued)**

**09:15–10:00 Session 2**

09:15–10:00 *Invited Talk 2: How (and Why) Online Dating Experiences Differ across American Cities*  
Elizabeth E. Bruch

**10:00–11:00 Lunch and Birds of a Feather Meetings**

**11:00–14:35 Session 3**

11:00–11:45 *Invited Talk 3: Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech*  
Jesse Shapiro

11:45–13:00 *Paper Session and Discussion 2*  
All papers

13:00–13:45 *Invited Talk 4: Persuasion, Bias, and Choice? Building Socially-aware Language Technologies*  
Diyi Yang

13:45–14:30 *Panel Discussion: NLP+CSS in Industry*  
Kristin Althenburger, Jason Baldrige, Glen Coppersmith, Emre Kiciman, and Umashanthi Pavalanathan