Brendan O'Connor

Associate Professor
College of Information and Computer Sciences
University of Massachusetts Amherst
140 Governors Drive, Amherst, MA, 01003, USA
brenocon@cs.umass.edu http://brenocon.com
Last updated Oct 2022

Education

Ph.D. in Machine Learning, 2014, Carnegie Mellon University	2009 – 2014
M.S. and B.S. in Symbolic Systems, 2006, Stanford University	2001 – 2006

Academic Positions

Associate Professor, Computer Science, University of Massachusetts Amherst	2020 –
Assistant Professor, Computer Science, University of Massachusetts Amherst	2014 – 2020
Visiting Fellow, Institute for Quantitative Social Science, Harvard University (Cambridge, MA)	Spring 2013
Graduate Research Assistant, Carnegie Mellon University (Pittsburgh, PA)	2009 – 2014

Industry Positions

Research Intern, Data Science group, Facebook, Inc. (Palo Alto, CA)	Summer 2010
Consulting Scientist, Crowdflower / Dolores Labs, Inc. (San Francisco, CA)	2008 – 2009
Engineer, Powerset, Inc. (San Francisco, CA)	2006 – 2008

Awards

- NSF CAREER Award, 2019.
- Facebook Robust Deep Learning for NLP Award, 2019.
- Best Paper Award at the 3rd Workshop on Noisy User-generated Text. 2017.
- Google Faculty Research Award, 2016.
- Finalist, Facebook Fellowship Program. 2010 and 2011.
- K. Jon Barwise Award for Distinguished Contribution to the Symbolic Systems Program. 2005.
- National AP Scholar, AP State Scholar, National Merit Scholar Finalist. 2001.

Refereed Main Conference Papers and Journal Articles

ACL, NAACL, EMNLP, CIKM, ICDM, ICTIR, WWW, and ICWSM are CS-style refereed conferences that publish peer reviewed, archival proceedings. Authors submit a paper, which is peer reviewed, then either accepted or rejected for publication in the conference proceedings. There is typically no revision cycle. Acceptance rates are typically 10-40%. This section only contatins main conference papers, excluding refereed workshop papers, since the conventional CS research view is that workshop papers typically contain smaller contributions, or work that is in progress.

Corpus-Level Evaluation for Event QA: The IndiaPoliceEvents Corpus Covering the 2002 Gujarat Violence.
Andrew Halterman, Katherine A. Keith, Sheikh Muhammad Sarwar, and Brendan O'Connor.
Findings of the Association for Computational Linguistics: ACL 2021.

How white is the global elite? An analysis of race, gender and network structure.
Kevin L. Young, Seth K. Goldman, Brendan O'Connor, and Tuugi Chuluun.
Global Networks. 2020.

Text and Causal Inference: A Review of Using Text to Remove Confounding from Causal Estimates.
Katherine A. Keith, David Jensen, and Brendan O'Connor.
Proceedings of ACL 2020 (Annual Meeting of the Association for Computational Linguistics). (full conference paper)

Investigating Sports Commentator Bias within a Large Corpus of American Football Broadcasts.
Jack Merullo, Luke Yeh, Abram Handler, Alvin Grissom II, Brendan O'Connor, Mohit Iyyer.
Proceedings of EMNLP 2019 (The 2019 Conference on Empirical Methods in Natural Language Processing). (full conference paper)

Query-focused Sentence Compression in Linear Time.
Abram Handler, Brendan O'Connor
Proceedings of EMNLP 2019 (Conference on Empirical Methods in Natural Language Processing). (short conference paper)

Uncertainty-aware generative models for inferring document class prevalence.
Katherine Keith, Brendan O'Connor.
Proceedings of EMNLP 2018 (Conference on Empirical Methods in Natural Language Processing). (full conference paper)

Understanding the Representational Power of Neural Retrieval Models Using NLP Tasks.
Daniel Cohen, Brendan O'Connor, W. Bruce Croft.
Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval. (full conference paper)

Exploring Diversification In Non-factoid Question Answering.
Lakshmi Vikraman, W. Bruce Croft, Brendan O'Connor.
Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval. (short conference paper)

A Neural Passage Model for Ad-hoc Document Retrieval.
Qingyao Ai, Brendan O'Connor, W. Bruce Croft.
European Conference on Information Retrieval. 2018. (conference paper)

Monte Carlo Syntax Marginals for Exploring and Using Dependency Parses.
Katherine Keith, Su Lin Blodgett, Brendan O'Connor.
Proceedings of NAACL 2018 (Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies). (full conference paper)

Relational Summarization for Corpus Analysis.
Abram Handler, Brendan O'Connor.
Proceedings of NAACL 2018 (Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies). (full conference paper)

Twitter Universal Dependency Parsing for African-American and Mainstream American English.
Su Lin Blodgett, Johnny Tian-Zheng Wei, Brendan O'Connor.
Proceedings of ACL 2018 (Annual Meeting of the Association for Computational Linguistics). (full conference paper)

A Probabilistic Approach for Learning with Label Proportions Applied to the US Presidential Election.
Tao Sun, Brendan O'Connor, Daniel Sheldon.
Proceedings of ICDM 2017. Acceptance rate 9% (72/778). (full conference paper)

Identifying civilians killed by police with distantly supervised entity-event extraction.
Katherine Keith, Abram Handler, Michael Pinkham, Cara Magliozzi, Joshua McDuffie, Brendan O'Connor.
Proceedings of EMNLP 2017 (Conference on Empirical Methods on Natural Language Processing). Acceptance rate 26% (216/836). (full conference paper)
Also presented at the *Data Science + Journalism workshop at KDD*, August 2017, and *Bloomberg Data for Good Exchange*, September 2017.

Learning to Extract Events from Knowledge Base Revisions.
Alexander Konovalov, Benjamin Strauss, Alan Ritter, Brendan O'Connor.
Proceedings of WWW 2017.
Acceptance rate 16% (164/996). (full conference paper)

Demographic Dialectal Variation in Social Media: A Case Study of African-American English.
Su Lin Blodgett, Lisa Green, Brendan O'Connor.
Proceedings of EMNLP 2016 (Conference on Empirical Methods on Natural Language Processing). Acceptance
rate 24%. (full conference paper)

Improving Entity Ranking for Keyword Queries.
John Foley, Brendan O'Connor, James Allan.
Proceedings of CIKM 2016 (25th ACM International Conference on Information and Knowledge Management).
Acceptance rate 24% (55/234 short papers). (short conference paper)

Posterior calibration and exploratory analysis for natural language processing models.
Khanh Nguyen, Brendan O'Connor.
Proceedings of EMNLP 2015 (Conference on Empirical Methods on Natural Language Processing). Acceptance
rate 24% (312/1315). (full conference paper)

Diffusion of Lexical Change in Social Media.
Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, Eric P. Xing.
PLoS ONE, PONE-D-14-28899R2, 2014. (journal paper)

Learning to Extract International Relations from Political Context.
Brendan O'Connor, Brandon Stewart, Noah A. Smith.
*Proceedings of ACL 2013 (Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics
(Volume 1: Long Papers)).* Acceptance rate 26% (174/664). (full conference paper)

Learning Latent Personae of Film Characters.
David Bamman, Brendan O'Connor, Noah A. Smith.
*Proceedings of ACL 2013 (Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics
(Volume 1: Long Papers)).* Acceptance rate 26% (174/664). (full conference paper)

Improved Part-of-Speech Tagging for Online Conversational Text with Word Clusters.
Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider and Noah A. Smith.
*Proceedings of NAACL-HLT 2013 (Proceedings of the 2013 Conference of the North American Chapter of the
Association for Computational Linguistics: Human Language Technologies).* Acceptance rate 30% (88/293). (full
conference paper). Longer tech report version: CMU-ML-12-107.

Censorship and Deletion Practices in Chinese Social Media.
David Bamman, Brendan O'Connor, Noah A. Smith.
First Monday 17.3, March 2012. (journal article)

Predicting a Scientific Community's Response to an Article.
Dani Yogatama, Michael Heilman, Brendan O'Connor, Chris Dyer, Bryan R. Routledge, Noah A. Smith.
Proceedings of EMNLP 2011 (Conference on Empirical Methods on Natural Language Processing). Acceptance
rate 24% (149/628). (full conference paper)

Part-of-Speech Tagging for Twitter: Annotations, Features, Experiments.
Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael
Heilman, Dani Yogatama, Jeff Flanigan, Noah A. Smith, 2011.
Proceedings of ACL 2011 (Proceedings of the Annual Meeting of the Association for Computational Linguistics).
Acceptance rate 26% (164/634). (short conference paper)

A Latent Variable Model for Geographic Lexical Variation.

Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, Eric P. Xing.
Proceedings of EMNLP 2010 (Conference on Empirical Methods on Natural Language Processing). Acceptance rate 25% (125/500). (full conference paper)

From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.
Brendan O'Connor, Ramnath Balasubramanyan, Bryan Routledge, Noah A. Smith.
Proceedings of ICWSM 2010 (Fourth International AAAI Conference on Weblogs and Social Media). Acceptance rate 19%. (full conference paper)

Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks.
Rion Snow, Brendan O'Connor, Daniel Jurafsky, Andrew Y. Ng.
Proceedings of EMNLP 2008 (Conference on Empirical Methods in Natural Language Processing). Acceptance rate 30% (116/385). (full conference paper)

Refereed Workshop and Demonstration Papers

These are peer reviewed, though typically less selective than main conference papers at CS venues. Some are accompanied by publication in conference proceedings (as is the case for ACL/EMNLP-affiliated workshops) while some are not (as is the case for NeurIPS -affiliated workshops and social science-style meetings).

Uncertainty over Uncertainty: Investigating the Assumptions, Annotations, and Text Measurements of Economic Policy Uncertainty.
Katherine A. Keith, Christoph Teichmann, Brendan O'Connor, and Edgar Meij.
NLP+CSS workshop at EMNLP 2020 *(Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science).*

Analyzing Gender Bias within Narrative Tropes.
Dhruvil Gala, Mohammad Omar Khursheed, Hannah Lerner, Brendan O'Connor, and Mohit Iyyer.
NLP+CSS workshop at EMNLP 2020 *(Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science).*

Summarizing Relationships for Interactive Concept Map Browsers.
Abram Handler, Premkumar Ganeshkumar, Brendan O'Connor, Mohamed AlTantawy.
Proceedings of the 2nd Workshop on New Frontiers in Summarization at EMNLP 2019.

Evaluating Syntactic Properties of Seq2seq Output with a Broad Coverage HPSG: A Case Study on Machine Translation.
Johnny Tian-Zheng Wei, Khiem Pham, Brian Dillon, Brendan O'Connor.
Proceedings of the Workshop on Analyzing and interpreting neural networks for NLP (BlackBoxNLP) at EMNLP 2018. Also arXiv:1809.02035.

Exploring Summary-Expanded Entity Embeddings for Entity Retrieval.
Shahrzad Naseri, John Foley, James Allan, Brendan O'Connor.
The EntitY REtrieval (EYRE) Workshop at CIKM 2018.

Racial Disparity in Natural Language Processing: A Case Study of Social Media African-American English.
Su Lin Blodgett, Brendan O'Connor.
Fairness, Accountability, and Transparency in Machine Learning (FAT/ML) (workshop at KDD 2017). Also arXiv:1707.00061.

A Dataset and Classifier for Recognizing Social Media English.
Su Lin Blodgett, Johnny Tian-Zheng Wei, Brendan O'Connor.
3rd Workshop on Noisy User-generated Text (WNUT) at EMNLP 2017.
Winner of Best Paper Award.

Rookie: A unique approach for exploring news archives.
Abram Handler, Brendan O'Connor.

Workshop on Data Science + Journalism at KDD 2017.

Bag of what? Simple noun phrase extraction for corpus analysis.
Abram Handler, Matt Denny, Hanna Wallach, Brendan O'Connor.
NLP+CSS 2016 – Workshop on Natural Language Processing and Computational Social Science (workshop at EMNLP 2016).

Visualizing textual models with in-text and word-as-pixel highlighting.
Abram Handler, Su Lin Blodgett, Brendan O'Connor.
WHI 2016 – Workshop on Human Interpretability in Machine Learning (workshop at ICML 2016).

Challenges of Visualizing Differentially Private Data.
Dan Zhang, Michael Hay, Gerome Miklau, Brendan O'Connor.
TPDP 2016 – Theory and Practice of Differential Privacy (workshop at ICML 2016). Selected for oral presentation.

MiTextExplorer: Linked brushing and mutual information for exploratory text data analysis.
Brendan O'Connor.
ACL Workshop on Interactive Language Learning, Visualization, and Interfaces, June 2014. Selected for oral presentation. (Also presented at *New Directions in Analyzing Text as Data*, October 2014.)

A framework for (under) specifying dependency syntax without overloading annotators.
Nathan Schneider, Brendan O'Connor, Naomi Saphra, David Bamman, Manaal Faruqui, Noah A. Smith, Chris Dyer, Jason Baldridge.
Linguistic Annotation Workshop, 2013. Also arXiv:1306.2091.

Mapping the geographical diffusion of new words.
Jacob Eisenstein, Brendan O'Connor, Noah A. Smith, and Eric P. Xing.
Poster at *NIPS 2012 Workshop on Social Network and Social Media Analysis*. Also arXiv:1210.5268.

Computational Text Analysis for Social Science: Model Assumptions and Complexity.
Brendan O'Connor, David Bamman, Noah A. Smith
NIPS 2011 Second Workshop on Comptuational Social Science and the Wisdom of Crowds.

A Mixture Model of Demographic Lexical Variation.
Brendan O'Connor, Jacob Eisenstein, Eric P. Xing, Noah A. Smith.
NIPS 2010 Workshop on Machine Learning for Social Computing.

TweetMotif: Exploratory Search and Topic Summarization for Twitter.
Brendan O'Connor, Michel Krieger, David Ahn.
ICWSM 2010 demonstration paper. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media.*

Groupware for Groups: Problem-Driven Design in Deme.
Todd Davies, Brendan O'Connor, Alex Cochran, Andrew Parker.
CHI 2005 Beyond Threaded Conversation Workshop.

Other Publications (Proceedings, Book Chapters, Book Review)

Su Lin Blodgett, Michael Madaio, Brendan O'Connor, Hanna Wallach, Qian Yang (ed).
Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing, at EACL 2021, online.

David Bamman, Dirk Hovy, David Jurgens, Brendan O'Connor, and Svitlana Volkova (ed).
Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science, at EMNLP 2020, online.

Preface: SCiL 2019 Editors' Note.
Gaja Jarosz, Max Nelson, Brendan O'Connor, Joe Pater.
Proceedings of the Society for Computation in Linguistics (SCiL) 2019, at LSA.

Gaja Jarosz, Brendan O'Connor, Joe Pater (ed).
Proceedings of the Society for Computation in Linguistics (SCiL) 2019, at LSA.

Preface: SCiL 2018 Editors' Note.
Gaja Jarosz, Brendan O'Connor, Joe Pater.
Proceedings of the Society for Computation in Linguistics (SCiL) 2018, at LSA.

Gaja Jarosz, Brendan O'Connor, Joe Pater (ed).
Proceedings of the Society for Computation in Linguistics (SCiL) 2018, at LSA.

Dirk Hovy, Svitlana Volkova, David Bamman, David Jurgens, Brendan O'Connor, Oren Tsur, A Seza Doğruöz.
Proceedings of the Second Workshop on NLP and Computational Social Science, at ACL 2017.

David Bamman, A Seza Doğruöz, Jacob Eisenstein, Dirk Hovy, David Jurgens, Brendan O'Connor, Alice Oh, Oren Tsur, Svitlana Volkova.
Proceedings of the First Workshop on NLP and Computational Social Science, at EMNLP 2016.

Superficial Data Analysis: Exploring Millions of Social Stereotypes.
Brendan O'Connor, Lukas Biewald. In *Beautiful Data: The Stories Behind Elegant Data Solutions*, ed. Toby Segaran and Jeff Hammerbacher. O'Reilly Media. 2009.

An Online Environment for Democratic Deliberation: Motivations, Principles, and Design.
Todd Davies, Brendan O'Connor, Alex Angiolillo Cochran, Andrew Parker, Jonathan Effrat, Benjamin Newman, Aaron Tam. In Todd Davies and Seeta Pena Gangadharan (ed.), Online Deliberation: Design, Research, and Practice, CSLI Publications/University of Chicago Press, 2009.

Book review of Donald Knuth, "Things a Computer Scientist Rarely Talks About."
Brendan O'Connor. In Raul Rojas (editor), "Reviews." IEEE Annals of the History of Computing, Oct-Dec 2004.

Conference Presentations or Papers with Abstract-only Submissions

These are abstract-only submissions accepted for presentation at social science conferences; they are typically associated with a non-published working paper shared with a few conference attendees. For Keith et al. 2018, Handler et al. 2016, and O'Connor 2014, the abstracts/presentations are associated with NLP papers listed in the previous workshops section.

Challenges in Detecting Null Relativizers in African American Language for Sociolinguistic and Psycholinguistic Applications.
Anissa Neal, Brendan O'Connor, and Lisa Green.
Presented at *UnImplicit: The First Workshop on Understanding Implicit and Underspecified Language*, at ACL-IJNLP 2021.

Deleted: the causal effects of online Reddit moderation on user and community behavior.
Katherine Keith, Brendan O'Connor.
Presented at *New Directions in Text as Data*, 2019.

Natural Language Processing Methods to Learn about Social Determinants of Health from Electronic Health Records.
David L. Chin, Vishal A. Shetty, Su Lin Blodgett, and Brendan T. O'Connor.
Presented at *AcademyHealth Annual Research Meeting*, 2019.

Uncertainty-aware generative models for inferring document class prevalence.
Katherine Keith, Brendan O'Connor.

Presented at *New Directions in Text as Data*, 2018.

A Little Bit of NLP Goes A Long Way: Adding Phrases to the Term-Document Matrix using Finite-State Shallow Parsing.
Abram Handler, Matthew J. Denny, Hanna Wallach, Brendan O'Connor.
Presented at *Seventh Annual New Directions in Analyzing Text as Data,* 2016. Acceptance rate 16% (19/120).

A Little Bit of NLP Goes A Long Way: Finding Meaning in Legislative Texts with Phrase Extraction.
Matthew J. Denny, Brendan O'Connor, Hanna Wallach.
Presented at *Midwest Political Science Association,* 2015.

MiTextExplorer: Linked brushing and mutual information for exploratory text data analysis.
Brendan O'Connor.
Presented at *New Directions in Analyzing Text as Data*, October 2014. Acceptance rate 29% (18/63). (See also ACL WS 2014 in earlier section.)

Theses and Dissertations

Statistical Text Analysis for Social Science.
Ph.D. dissertation, Carnegie Mellon University, August 2014.
Committee: Noah Smith (CMU, Language Technologies Institute), Tom Mitchell (CMU, Machine Learning Department), Cosma Shalizi (CMU, Statistics Department), Gary King (Harvard, Government Department).

Biased evidence assimilation under bounded Bayesian rationality.
M.S. thesis, Stanford University, 2006.
Readers: Jonathan Bendor (Political Science), James McClelland (Psychology).

The social evolution of human cooperation via group competition and conflict.
B.S. honors thesis, Stanford University, 2006.
Readers: Noah Mark (Sociology), James Fearon (Political Science).

Other Non-refereed Documents and Preprints
This section does not include arxiv preprints that are also published in a peer-reviewed venue.

Human acceptability judgements for extractive sentence compression.
Abram Handler, Brian Dillon, Brendan O'Connor.
arXiv:1902.00489, February 2019.

Annotation Guidelines for Twitter Part-of-Speech Tagging (Version 0.3).
Kevin Gimpel, Nathan Schneider, Brendan O'Connor.
March 2013.

ARKref: a rule-based coreference resolution system.
Brendan O'Connor, Michael Heilman.
arXiv:1310.1975, October 2013.

Learning Frames from Text with an Unsupervised Latent Variable Model.
Brendan O'Connor.
arXiv:1307.7382, Data Analysis Project report, Machine Learning Department, CMU. July 2013.

Music Genre Classification with the Million Song Dataset.
Dawen Liang, Haijie Gu, Brendan O'Connor.
Course project, December 2011.

Grants – Government (NSF)

- Kristine Yu (PI), Lisa Green (co-PI), Meghan Armstrong-Abrami (co-PI), and Brendan O'Connor (co-PI). National Science Foundation. "Understanding variation in African American Language: Corpus and prosodic fieldwork perspectives." 2021-2024. NSF #2042939 (SBE BCS). $434,027.
- Brendan O'Connor (PI) and Mohit Iyyer (co-PI). National Science Foundation. "CCRI: Planning for the Development of a Platform to Support Multilingual and Multidomain Coreference Annotation in Natural Language Processing Research." 2019-2020. NSF #1925548 (CISE CNS). $99,998.
- Brendan O'Connor (PI). National Science Foundation. "CAREER: Social Aggregate Measurement from Text." 2019-2023. NSF #1845576 (CISE IIS). 2019-2024. $545,963.
- Brendan O'Connor (PI). National Science Foundation. "III: Small: Collaborative Research: Building subjective knowledge bases by modeling viewpoints." 2018-2021. NSF #1814955 (CISE IIS). $249,978 (out of $515,978 collaborative grant with David Bamman, UC Berkeley).
- Joseph Pater (PI) and Brendan O'Connor (co-PI). National Science Foundation. "Conference: Perceptrons and Syntactic Structures at 60: Computational Modeling of Language." NSF #1651142 (SBE BCS). 2017-2019. $24,184.

Grants – Industry

- Brendan O'Connor, Mohit Iyyer. Facebook, Inc. Facebook Robust Deep Learning for Natural Language Processing Research Award. "Robust and fair dialectal NLP via unsupervised social media LM transfer." 2019-2020. $76,905.
- Brendan O'Connor, Abram Handler, Brian Dillon. Figure Eight, Inc. AI for Everyone Challenge. "Crowdsourced tree edit supervision for flexible sentence summarization and corpus exploration." 2018-2020. $25,000.
- Brendan O'Connor. Google, Inc. Google Faculty Research Award. "Eliminating racial bias in parsing of online language." 2017-2018. $50,538.
- Brendan O'Connor. Sanfkofa, Inc. "Discovery of demographically grounded cultural trends from social media text." 2016. University of Massachusetts Innovation Institute Scoping Project. $20,000.

Grants – Internal

- Kevin Young (Political Science), Seth Goldman (Communication), Brendan O'Connor (Computer Science). Faculty Research Grant/Healey Endowment Grant (FRG/HEG). "Networks of Organizational Leadership: Mapping race and gender diversity change among elites." 2019-2020. $8,469.
- Kevin Young (Political Science), Seth Goldman (Communication), Brendan O'Connor (Computer Science). Institute of Diversity Sciences seed award. "Networks of Organizational Leadership: Measuring Race, Gender, and Color Diversity among Elites." 2019-2020. $9,059.
- Kevin Young (Political Science), Seth Goldman (Communication), Brendan O'Connor (Computer Science). College of Social and Behavioral Sciences Research Support Grant. "Changing Diversity in the Power Elite: Networks, Evolution and Constraint." 2018-2019. $17,108.
- Meghan Armstrong-Abrami (Hispanic Linguistics), Lisa Green (Linguistics), Brendan O'Connor (Computer Science), Kristine Yu (Linguistics). Institute of Diversity Sciences seed award. "Discovering African American English Speech Melodies." 2018-2019. $12,000.

Selected Press Coverage and Interviews

- Diaries of Social Research, 2021. Podcast interview. (With Brandon Stewart, Katherine Keith, and Lucy Li.)
- NLP Highlights, 2020, #103. Podcast interview. "Processing Language in Social Media." (With Waleed Ammar and Pradeep Dasigi.)
- New Scientist, Sept. 22, 2017. Matt Reynolds. "This AI reads the news to keep tabs on US police shootings."
- MIT Tech Review, August 2017. Will Knight. "AI Programs Are Learning to Exclude Some African-American Voices."
- Inverse, Sept. 22, 2016. Tonya Riley. "A.I. Doesn't Get Black Twitter."
- New Yorker, Feb. 9, 2015. Reeves Wiedeman. "Talk to me."

- Time, Sep. 9, 2013. Katy Steinmetz. "What Twitter Says to Linguists."
- BBC, Mar. 9, 2012. "China's social networks hit by censorship, says study."
- New York Times, Oct. 29, 2011. Ben Zimmer. "Twitterology: A New Science?"
- Wall Street Journal, Feb. 12, 2011. Christopher Shea. "Week in Ideas."
- BBC Radio 5, Oct. 2010. "Neighbours, sentiments and sketched out tweets."

Teaching and Advising

<u>University of Massachusetts Amherst</u>
- Instructor, CS 485/490A, Applications of Natural Language Processing. Fall 2020, Fall 2021.
- Instructor, CS 685/690N, Advanced Natural Language Processing. Spring 2017, Spring 2018, Spring 2020, Spring 2021.
- Co-instructor (with Mohit Iyyer), CS 690D, Deep Learning for Natural Language Processing. Spring 2019.
- Instructor, CS 688, Probabilistic Graphical Models. Spring 2016.
- Instructor, CS 589, Machine Learning. Fall 2018.
- Instructor, CS 585, Introduction to Natural Language Processing. Fall 2014, Fall 2015, Fall 2016, Fall 2017.
- Instructor, CS H585, Honors Colloquium for Introduction to Natural Language Processing. Fall 2016, Fall 2018.
- Instructor, CS 691MA, Social Media Analysis and Computational Social Science. Spring 2015.
- Supervised six undergraduate honors theses and projects (5 Computer Science students, 1 Linguistics student), resulting in conference and workshop publications (Nguyen and O'Connor EMNLP 2015, Wei et al. BlackBoxNLP 2018).
- Supervised seven Research Experience for Undergraduates participants in CICS. Summers of 2015, 2016, 2017, and 2019. Students from Columbia, University of Kentucky, and University of Massachusetts Amherst. Four were supervised within the NSF REUMass site REU.
- Co-advised three masters project theses.
- Instructor, Summer Dialect Research Project (SDRP), at the Center for the Study of African American Language at UMass Amherst (with director Lisa Green). Teaching and advising undergraduate students from historically black colleges and universities. Supervised two undergraduates' projects on NLP and African-American English. Summer 2019.
- Ph.D. student alumni
  - Su Lin Blodgett, graduated 2020.
  - Abram Handler, graduated 2021.
  - Katherine Keith, graduated 2021.
- Current Ph.D. students
  - Ankita Gupta, 2021-.
  - Tessa Massis, 2021-.
  - Erica Cai, 2022-.
- Co-advised 11 Ph.D. student synthesis projects: 3 for my advisees, and 8 others as a co-advisor.
- Member – Ph.D. dissertation committees (not up to date)
  - Pat Verga (Computer Science), 2019.
  - Emma Strubell (Computer Science), 2018-2019.
  - Myungha Jang (Computer Science), 2018-2019.
  - John Foley (Computer Science), 2018-2019.
  - Coral Hughto (Linguistics), 2018-2019.
  - John Lalor (Computer Science), 2018-2019.
  - Hsin-Lun Huang (Linguistics), 2017-2018.
  - Jiepu Jiang (Computer Science), 2017.

<u>Carnegie Mellon University (as graduate student)</u>
- Supervised undergraduate research assistant. Summer 2012. Resulted in conference publication (Owoputi et al. NAACL 2013), and popular open-source software package (CMU ARK Twitter POS Tagger: http://www.ark.cs.cmu.edu/TweetNLP/)

- Teaching Assistant, 10-710, Structured Prediction (Cohen, Smith). 2011.
- Teaching Assistant, 10-601, Machine Learning (Mitchell, Bar-Joseph). 2012.

Stanford University (as undergraduate student)
- Student Initiated Course – Developed and taught seminar on Douglas Hofstadter's *Gödel, Escher, Bach*.
- Section Leader (TA), CS 106A, Introduction to Programming.
- Advising Fellow – Undergraduate student advisor for the interdisciplinary Symbolic Systems Program. Academic peer advising, organizing speakers and campus events, and giving input to curriculum decisions.

Internal Committee Service

- Annual faculty review personnel committee, 2015, 2020.
- Co-chair, informatics faculty hiring committee, 2019. Reviewed applications, organized and hosted visits, gathered and summarized feedback, made hiring recommendations to the faculty, hosted second visits, for the new informatics initiative in UMass CICS.
- Computational linguistics major cross-college committee, 2016-present. Working with the Department of Linguistics, designed a new interdisciplinary major combining Computer Science and Linguistics, as part of CICS' CS+X initiative. Helped draft materials for the official major proposal, advocated to and gathered feedback from CICS faculty, and managed administrative issues with current CL students, especially course admissions.
- Informatics faculty hiring committee, 2018. Reviewed applications, organized and hosted faculty job candidate visits for the new informatics initiative in UMass CICS.
- Graduate awards committee, 2019.
- Informatics program committee, 2017.
- Graduate program committee, 2017. Reviewing of doctoral student progress and researching and proposing solutions to degree requirement issues.
- Cognitive Science Initiative steering committee, 2015-2019.
- Graduate admissions committee, 2015, 2016 (doctoral), 2018 (masters), 2020 (doctoral). Reviewed applications for CICS masters and doctoral programs; managed general reviewing and diversity among application reviews.
- Data science faculty hiring committee, 2016.

Academic Service – Discussant
A paper discussant reviews preprints of several papers for a conference session, and gives a presentation summarizing and/or critiquing them.

Discussant at *New Directions in Analyzing Text as Data*, October 2017, Princeton, NJ.

Academic Service – Senior Program Committee and Conference Organization
NLP conferences use the term "Program Committee" broadly, to refer to most of the reviewer pool, while some other communities, such as systems-style conferences, use the term to specify a smaller group. I'm excluding the broad use of the term "PC" in this section. I am including workshop and area chairing duties as a rough equivalent of a "(Senior) PC" as the term is used outside NLP. Note AC roles before 2016 combine duties that are not split between ACs and SACs.

- Member, Ethics Committee, EMNLP 2021.
- Co-organizer, First Workshop on Bridging Human–Computer Interaction and Natural Language Processing, at EACL 2021. https://sites.google.com/view/hciandnlp
- Area chair, Theme Track: NLP for Social Good, at ACL-IJCNLP 2021.
- Area chair, Computational Social Science and Cultural Analytics track, at NAACL-HLT 2021.
- Co-organizer, Workshops on Natural Language Processing and Computational Social Science (NLP+CSS), at WebSci 2016, EMNLP 2016, ACL 2017, and EMNLP 2020. Helped found and organize interdisciplinary workshop, bringing together social scientists and computer scientists. https://sites.google.com/site/nlpandcss/home
- Co-organizer, Society for Computation in Linguistics, 2018 and 2019. https://blogs.umass.edu/scil/

- Workshop co-chair, Association for Computational Linguistics (ACL) 2018. (Co-chair: Eva Maria Vecchi) Responsible for soliciting and reviewing workshop applications across several NLP conferences, and coordinating resulting workshops at ACL.
- Senior program committee member, International Conference on Weblogs and Social Media, 2018-2019. Responsible for assigning and supervising reviews for paper submissions.
- Area chair, Computational Social Science and Social Media track, at EACL 2020.
- Area chair, NLP for Social Media and Social Sciences track, at NAACL-HLT 2015 (Conference: North American Chapter for the Association for Computational Linguistics / Human Language Technologies). (Co-chair: Philip Resnik.) Responsible for assigning and supervising reviews for several dozen paper submissions, and selection recommendations.
- Area chair, NLP for the Web and Computational Social Science track, for EMNLP 2015 (Conference: Empirical Methods in Natural Language Processing). (Co-chair: Alice Oh.) Responsible for assigning and supervising reviews for several dozen paper submissions, and selection recommendations.

Academic Service – Reviewing

For conferences and workshops, years correspond to the year of the conference event. These lists exclude years where I served as organizer, on senior PC, or as area chair (above).

Grant reviewing
- Panelist, National Science Foundation. 2016 (CISE/SBE), 2021 (CISE), 2022 (CISE).
- Ad-hoc reviewer, National Science Foundation. 2014 (SBE), 2015 (SBE), 2020 (CISE).
- Reviewer, Marsden Fund, Royal Society of New Zealand. 2018.
- Reviewer, National Council of Science and Technology of Chile. 2018.
- Reviewer, MIDAS, University of Michigan. 2016.

Natural language processing and computational linguistics
- ACL (Conference: Association for Computational Linguistics). 2014, 2016, 2017.
- EMNLP (Conference: Empirical Methods in Natural Language Processing). 2011, 2013, 2015, 2016, 2018, 2019.
- NAACL-HLT (Conference: North American Chapter for the Association for Computational Linguistics / Human Language Technologies). 2013, 2015, 2018, 2019.
- TACL (Journal: Transactions of the Association for Computational Linguistics). 2013-2018. Joined "standing reviewer team" in Fall 2014.
- AKBC (Conference: Automated Knowledge Base Construction). 2021.
- CoNLL (The SIGNLL Conference on Computational Natural Language Learning). 2018, 2019, 2021.
- EACL (Conference: European Chapter of the Association for Computational Linguistics). 2013, 2014, 2017.
- Semeval (task proposal review). 2016.
- *SEM (Joint Conference on Lexical and Computational Semantics). 2013.
- IJCNLP (International Joint Conference on Natural Language Processing). 2011.

Machine learning
- NeurIPS (Conference on Neural Information Processing Systems). 2015, 2019.
- AISTATS (International Conference on Artificial Intelligence and Statistics). 2016.
- ICML (International Conference on Machine Learning). 2011, 2013, 2014, 2015.
- KDD (ACM SIGKDD Conference on Knowledge Discovery and Data Mining). 2015.
- Neurocomputing (journal). 2014.
- Machine Learning Journal. 2012.

Statistics
- Bayesian Analysis (journal). 2016.
- JASA (Journal of the American Statistical Association). 2014.
- Biostatistics (journal). 2013.

Other CS/AI/HCI venues, journals, and academic press

- VIS (IEEE Visualization Conference). 2021.
- COMPASS (ACM Conference on Computing and Sustainable Societies). 2018.
- JAIR (Journal of Artificial Intelligence Research). 2014, 2018.
- InfoVis (Conference: IEEE Information Visualization, part of IEEE VIS). 2017.
- TIIS (Transactions on Interactive Intelligent Systems). 2017.
- MIT Press. 2017.
- IUI (ACM International Conference on Intelligent User Interfaces). 2017.
- ICWSM (International Conference on Weblogs and Social Media). 2011, 2014.
- CSCW (ACM Conference on Computer-Supported Cooperative Work and Social Computing). 2014.
- CHI (ACM SIGCHI Conference on Human Factors in Computing Systems). 2010, 2013.
- WWW (World Wide Web Conference). 2011, 2017.
- CACM (Communications of the ACM). 2011.

Conference workshops in CS (NLP, ML, etc..)
- Workshop on Social Sensing (SocialSens) at ICWSM. 2021.
- Natural Language Processing and Computational Social Science (NLP+CSS). 2019.
- AI for Social Good workshop at NeurIPS. 2019.
- Ethics in Natural Language Processing (EthNLP). 2017, 2018.
- Workshop on Noisy User-Generated Text (WNUT). 2017.
- ACL Workshop on NLP and Social Dynamics (NLPSD). 2014.
- ACL Workshop on Language Technologies and Computational Social Science (LACSS). 2014.
- Creating Speech and Language Data With Amazon's Mechanical Turk (at NAACL-HLT). 2010.
- Topic Models workshop at NeurIPS. 2013.
- Politics, Elections and Data (PLEAD) at CIKM. 2013.
- The 4th Human Computation Workshop (HCOMP). 2012.
- Computational Social Science and Wisdom of Crowds (CSS) workshop at NeurIPS. 2011.

Social science and general science journals
- PLOS ONE. 2013, 2015, 2016.
- Public Opinion Quarterly. 2013, 2014.
- Political Analysis. 2012.
- Social Science Computing Review. 2012.

Tutorials

- National Institutes of Health, Office of Behavioral and Social Sciences (OBSSR) Methodology Seminar. Rockville, MD. "Tutorial on natural language processing for social science investigation." August 2019.
- Tutorial on R. (With Michael Driscoll.) Amyris, Inc., Emeryville, CA. June 2009.

Invited Talks, Panels, Participation

- Talk, Allen Institute for Artificial Intelligence, Workshop on AI and Fairness. Seattle, WA. August 2019.
- Panelist, NSF Workshop on Computing, Data Science, and Access to Justice. Georgetown University. Washington, DC. June 2019.
- Talk, Johns Hopkins University, Center for Language and Speech Processing. Baltimore, MD. February 2019.
- Invited participant, Google, Credibility on the Web Workshop. New York, NY. October 2018.
- Invited participant, New York University Center for Data Science, Text Analysis Developers' Workshop. New York, NY. April 2018.
- Talk, Carnegie Mellon University, Language Technologies Institute (LTI Seminar). Pittsburgh, PA. March 2018.
- Talk, Yale University, The Data Science Workshop on Computational Social Science. New Haven, CT. October 2017.
- Talk, Google, Natural Language Understanding Workshop. New York, NY. July 2017.

- Talk, New York University Center for Data Science, NLP and Text as Data seminar. New York, NY. April 2017.
- Talk, Michigan State University, Social Science Data Analytics. East Lansing, MI. April 2017.
- Outreach: lecture and discussion at 4th Annual MLK Observance Dinner and Dialogue. Springfield, MA. January 2017.
- Talk, Northeastern University, Network Science Institute. Boston, MA. December 2016.
- Talk, WPI Computer Science Colloquium. Worcester Polytechnic Institute. Worcester, MA. September 2016.
- Talk, New England Machine Learning Day. Microsoft Research New England. Cambridge, MA. May 6, 2016.
- Talk, ISAT/DARPA What If: Machine Learning for Causal Inference Workshop. Cambridge, MA and Deadham, MA. February 11-12, 2016.
- Panelist, National Council on Measurement in Education, panel on "Fairness and Machine Learning for Educational Practice." Washington, DC. April 2016.
- Invited participant, Microsoft Research Faculty Summit, July 8-9, 2015.
- Talk, University of Edinburgh, Institute for Language, Cognition and Computation. Edinburgh, UK. September 2015. (Host: Shay Cohen)
- Talk, ACL 2015 Workshop on Noisy User-generated Text. Beijing, China. July 2015.
- Talk, University of Michigan, School of Information. March 2014. (Host: Qiaozhu Mei)
- Talk, New York University, Courant Institute and Center for Data Science. March 2014. (Host: Denis Zorin)
- Talk, Allen Institute for Artificial Intelligence. February 2014. (Host: Oren Etzioni)
- Talk, University of Washington, Information School. February 2014. (Host: Joshua Blumenstock)
- Talk, Microsoft Research New York City. February 2014. (Host: Sharad Goel)
- Talk, University of Massachusetts Amherst, School of Computer Science. February 2014. (Host: Andrew McCallum)
- Talk, Cornell University, Information Science Department. February 2014. (Host: David Mimno)
- Talk, University of Chicago, Toyota Technological Institute at Chicago. January 2014.
- Talk, University of Pennsylvania, Wharton School, Statistics Department. January 2014.
- Talk, University of Maryland, College Park, CLIP Colloquium (Computational Linguistics and Information Processing). "Text Analysis and Social Science: Learning to Extract International Relations from the News." October 2013. (Host: Philip Resnik, Linguistics and UMIACS)
- Talk, Northeastern University, Lazer Lab. "Social Text Data Analysis." March 2013. (Host: David Lazer, Political Science and Computer and Information Science)
- Talk, Columbia University, NLP group. February 2013.
- Talk, University of Chicago Computational Social Science Workshop (seminar series). "Social Text Data Analysis." November 2012. (Host: Forest Gregg, Sociology)
- Talk, UMass Amherst, Machine Learning and Friends Lunch and Computational Social Science seminar. "Text-as-Data Analysis." October 2012. (Host: Hanna Wallach, Computer Science)
- Talk, Penn State Political Science Department, New Faces in Political Methodology V workshop. "Corpus Analysis and Unsupervised Frame Learning from Text." April 2012. (Host: Burt Monroe, Political Science)
- Panelist, American Association for Public Opinion Research (AAPOR) conference, panel on "Survey Responses vs. Tweets: New Choices for Social Measurement." "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." May 2012. (Host: Frederick Conrad, Institute of Social Research, Univ. of Michigan)
- Talk, CMU Machine Learning Lunch. "Modeling Social Phenomena in Microblog Text." April 2011.
- Talk, "Dolores Labs and Amazon Mechanical Turk."
  - Stanford Human-Computer Interaction group. September 2008. (Host: Michel Krieger, Stanford/Meebo)
  - Stanford Natural Language Processing group. May 2008. (Host: Rion Snow, Stanford/Twitter)
  - Freebase / Metaweb Technologies, San Francisco, CA. May 2008. (Host: Brian Karlak, Metaweb)
  - Dataviz Salon SF, San Francisco, CA. February 2009. (Host: Michael Driscoll, Dataspora)

Software Projects and Artifacts

Phrasemachine. 2016.

Abram Handler, Matthew Denny, Brendan O'Connor.
http://slanglab.cs.umass.edu/phrasemachine/

Mutual Information Text Explorer. 2014.
Brendan O'Connor.
Open source, interactive analysis tool for text data and document covariates.
http://brenocon.com/mte/

CMU ARK Twitter NLP Tools. 2011-2013.
Tokenizer, part-of-speech tagger, and annotated data for Twitter.
Open source release (thousands of downloads)
http://www.ark.cs.cmu.edu/TweetNLP/

ARKref. 2011.
Brendan O'Connor, Michael Heilman.
Rule-based coreference system. Open source.
http://www.ark.cs.cmu.edu/ARKref/

TweetMotif.com. 2010.
Brendan O'Connor, Michel Krieger, David Ahn.
Search and summarization for Twitter. Open source. Tokenizer is in widespread use.

PalinSpeak.com. 2008.
Brendan O'Connor, Douglas Wilson.
Markov model chatbot as political satire.

Fleshmap.com. 2008.
(Primary project by Fernanda Viegas and Martin Wattenberg.)
Art project: Mechanical Turk data collection supporting the spatial visualization of human sexual preferences.
Press coverage: NPR

Patents

Efficient Storage and Retrieval of Posting Lists.
Chad Walters, Giovanni Lorenzo Thione, Barney Pell, Lukas Biewald, Brendan O'Connor.
US 8229970 B2 (Application US 12/201,079)
Filed August 29, 2008. Powerset, Inc. and Microsoft Corporation.

Emphasizing Search Results According to Conceptual Meaning.
Barney Pell, Scott Prevost, Giovanni Lorenzo Thione, Brendan O'Connor, Lukas Biewald.
US 8209321 B2 (Application US 12/201,504)
Filed August 29, 2008. Powerset, Inc. and Microsoft Corporation.

Browsing Knowledge on the Basis of Semantic Relations.
Franco Salvetti, Giovanni Lorenzo Thione, Richard Crouch, Lukas Biewald, Brendan O'Connor, Barney Pell.
US 8209321 B2 (Application US 12/201,978).
Filed August 29, 2008. Powerset, Inc. and Microsoft Corporation.