

Brendan O'Connor

Assistant 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 August 2016

Education

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

Academic Positions

Assistant Professor, Computer Science, University of Massachusetts Amherst Sept 2014 – present
Visiting Fellow, Institute for Quantitative Social Science, Harvard University (Cambridge, MA) Jan–Mar 2013
Graduate Research Assistant, Carnegie Mellon University (Pittsburgh, PA) Sept 2009 – Aug 2014

Industry Positions

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

Awards

- Finalist, Facebook Fellowship Program. 2010 and 2011.
- K. Jon Barwise Award for Distinguished Contribution to the Symbolic Systems Program. 2005.
- National AP Scholar (2001), AP State Scholar (2001), National Merit Scholar Finalist (2001), Qualifier for American Invitational Mathematics Examination (2000, 2002).

Selected Press Coverage of Research

- 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.”

Publications

The following three sections list refereed conference papers, journal articles, workshop papers, and book chapters. Non- or lightly-refereed works are listed later in this document.

Refereed Main Conference Papers and Journal Articles

Note: ACL, NAACL, EMNLP, CIKM and ICWSM are CS-style refereed conferences with a peer-reviewed, archival proceedings publication. Acceptance rates are typically 20-40%. Full papers are typically more selective than short papers. Refereed workshop papers are not in this section.

Demographic Dialectal Variation in Social Media: A Case Study of African-American English.

Su Lin Blodgett, Lisa Green, and 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, and James Allan.

Proceedings of CIKM 2016 (25th ACM International Conference on Information and Knowledge Management). Acceptance rate 29% (270/935). (short conference paper)

Posterior calibration and exploratory analysis for natural language processing models.

Khanh Nguyen and 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, and Eric P. Xing.

PLoS ONE, PONE-D-14-28899R2, forthcoming Nov 14, 2014. (journal paper)

Learning to Extract International Relations from Political Context.

Brendan O'Connor, Brandon Stewart, and 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, and 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, and 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, and 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 full conference papers at CS venues.

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, and Brendan O'Connor.
At *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 and Brendan O'Connor.
At *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. (Work 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, and 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 Computational 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.

Book Chapters

“Superficial Data Analysis: Exploring Millions of Social Stereotypes.”

Brendan O'Connor and 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 (editors), *Online Deliberation: Design, Research, and Practice*, CSLI Publications/University of Chicago Press, 2009.

Teaching and Advising

University of Massachusetts Amherst

- (Forthcoming) Instructor, CS 690N, Natural Language Processing. Spring 2017.
- Instructor, CS 688, Probabilistic Graphical Models. Spring 2016.
- Instructor, CS 585, Introduction to Natural Language Processing. Fall 2014, Fall 2015, Fall 2016.
- Instructor, CS 691MA, Social Media Analysis and Computational Social Science. Spring 2015.
- Supervising three PhD students, UMass Computer Science.
- Supervised two undergraduate honors theses, UMass Computer Science, resulting in conference publication (EMNLP 2015).
- Supervised three Research Experience for Undergraduates participants (Summer 2015, Summer 2016 in the NSF REUMass program).
- Supervised five other undergraduate independent studies (students from UMass CS, UMass Linguistics).

Carnegie Mellon University (graduate)

- Supervised undergraduate research assistant. 2012.
Resulted in conference publication, 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.

Corporate

- Tutorial on R. (With Michael Driscoll.) Amyris, Inc., Emeryville, CA. June 2009.

Stanford University (undergraduate)

- 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.

Academic Service – Program Committee and Organization

Co-organizer, Workshops on Natural Language Processing and Computational Social Science (NLP+CSS), at WebSci 2016 and EMNLP 2016.

Area co-chair (rough equiv. PC member in non-NLP conferences), NLP for Social Media and Social Sciences track, for NAACL-HLT 2015 (Conference: North American Chapter for the Association for Computational Linguistics / Human Language Technologies). (Co-chair: Philip Resnik)

Area co-chair (rough equiv. PC member in non-NLP conferences), NLP for the Web and Computational Social Science track, for EMNLP 2015 (Conference: Empirical Methods in Natural Language Processing). (Co-chair: Alice Oh)

Academic Service – Reviewing

For conferences and workshops, years correspond to the year of the conference event.

Grant reviewing

- Panelist, National Science Foundation. 2016. (CISE/IIS, SBE/BCS)
- Ad-hoc reviewer, National Science Foundation. 2014. (SBE/SES)

Natural language processing and computational linguistics

- Semeval (task proposal review). 2016.
- TACL (Journal: Transactions of the Association for Computational Linguistics). 2016, 2015, 2014, 2013. Joined “standing reviewer team” in Fall 2014.
- EMNLP (Conference: Empirical Methods in Natural Language Processing). 2016, 2015, 2013, 2011.
- ACL (Conference: Association for Computational Linguistics). 2016, 2014.
- NAACL-HLT (Conference: North American Chapter for the Association for Computational Linguistics / Human Language Technologies). 2015, 2013.
- EACL (Conference: European Chapter of the Association for Computational Linguistics). 2014, 2013.
- *SEM (Joint Conference on Lexical and Computational Semantics). 2013.
- IJCNLP (International Joint Conference on Natural Language Processing). 2011.

Machine Learning

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

Statistics

- JASA (Journal of the American Statistical Association). 2014.
- Biostatistics (journal). 2013.

AI, HCI, and general CS

- ICWSM (International Conference on Weblogs and Social Media). 2014, 2011.
- JAIR (Journal of Artificial Intelligence Research). 2014.
- CSCW (ACM Conference on Computer-Supported Cooperative Work and Social Computing). 2014.
- CHI (ACM SIGCHI Conference on Human Factors in Computing Systems). 2013, 2010.
- WWW (World Wide Web Conference). 2011.
- CACM (Communications of the ACM). 2011.

CS conference workshops

- ACL Workshop on NLP and Social Dynamics (NLPSD). 2014.
- ACL Workshop on Language Technologies and Computational Social Science (LACSS). 2014.
- NAACL-HLT workshop: Creating Speech and Language Data With Amazon’s Mechanical Turk. 2010.
- NIPS workshop: Topic Models workshop. 2013.
- CIKM workshop: Politics, Elections and Data (PLEAD). 2013.
- Workshop on Human Computation & Crowdsourcing (HCOMP). 2012.
- NIPS workshop: Computational Social Science and Wisdom of Crowds (CSS). 2011.

Social science and general science journals

- PLOS ONE. 2015, 2013.
- Public Opinion Quarterly. 2014, 2013.
- Political Analysis. 2012.

- Social Science Computing Review. 2012.

Invited Talks, Panels, Participation

- Talk, WPI Computer Science Colloquium. Worcester Polytechnic Institute. Sep, 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. Feb 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. Sep 2015. (Host: Shay Cohen)
- Talk, ACL 2015 Workshop on Noisy User-generated Text. Beijing, China. Jul 2015.
- Talk, University of Michigan, School of Information. Mar 2014. (Host: Qiaozhu Mei)
- Talk, New York University, Courant Institute and Center for Data Science. Mar 2014. (Host: Denis Zorin)
- Talk, Allen Institute for Artificial Intelligence. Feb 2014. (Host: Oren Etzioni)
- Talk, University of Washington, Information School. Feb 2014. (Host: Joshua Blumenstock)
- Talk, Microsoft Research New York City. Feb 2014. (Host: Sharad Goel)
- Talk, University of Massachusetts Amherst, School of Computer Science. Feb 2014. (Host: Andrew McCallum)
- Talk, Cornell University, Information Science Department. Feb 2014. (Host: David Mimno)
- Talk, University of Chicago, Toyota Technological Institute at Chicago. Jan 2014.
- Talk, University of Pennsylvania, Wharton School, Statistics Department. Jan 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.” Oct. 2013. (Host: Philip Resnik, Linguistics and UMIACS)
- Talk, Northeastern University, Lazer Lab. “Social Text Data Analysis.” Mar 2013. (Host: David Lazer, Political Science and Computer and Information Science)
- Talk, Columbia University, NLP group. Feb 2013.
- Talk, University of Chicago Computational Social Science Workshop (seminar series). “Social Text Data Analysis.” Nov 2012. (Host: Forest Gregg, Sociology)
- Talk, UMass Amherst, Machine Learning and Friends Lunch and Computational Social Science seminar. “Text-as-Data Analysis.” Oct 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.” Apr 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.” Apr 2011.
- Talk, “Dolores Labs and Amazon Mechanical Turk.”
 - Stanford Human-Computer Interaction group. Sep 2008. (Host: Michel Krieger, Stanford/Meebo)
 - Stanford Natural Language Processing group. May 2008. (Host: Rion Snow, Stanford/Twitter)
 - Freebase / Metaweb Technologies, Inc., San Francisco, CA. May 2008. (Host: Brian Karlak, Metaweb)
 - Dataviz Salon SF, San Francisco, CA. Feb 2009. (Host: Michael Driscoll, Dataspora)

Abstract-only Conference 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 (in the case of Denny et al. 2015). For Handler et al. 2016 and O’Connor 2014, the abstracts/presentations are associated with NLP workshop papers listed in the previous workshops section.

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.

To be 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, and 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.)

Non-refereed Documents

Annotation Guidelines for Twitter Part-of-Speech Tagging (Version 0.3)

Kevin Gimpel, Nathan Schneider, Brendan O'Connor.

Mar 2013.

ARKref: a rule-based coreference resolution system.

Brendan O'Connor and Michael Heilman.

arXiv:1310.1975, Oct 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, Dec 2011.

The social evolution of human cooperation via group competition and conflict.

B.S. honors thesis, Stanford University, 2006.

Readers: Noah Mark (Sociology) and James Fearon (Political Science).

Biased evidence assimilation under bounded Bayesian rationality.

M.S. thesis, Stanford University, 2006.

Readers: Jonathan Bendor (Political Science) and James McClelland (Psychology).

Dissertation

Statistical Text Analysis for Social Science. 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)

Book Review

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.

Software Projects and Artifacts

Mutual Information Text Explorer. 2014.
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 and Michael Heilman.
Rule-based coreference system. Open source.
<http://www.ark.cs.cmu.edu/ARKref/>

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

PalinSpeak.com. 2008.
Brendan O'Connor and 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.