

Brendan O'Connor

5817 Darlington Rd Apt 5, Pittsburgh PA 15217
brenocon@gmail.com <http://anyvall.org> cell: 650-823-9570
Last updated Feb. 2010

Education

Carnegie Mellon University, Pittsburgh, PA *Sept 2009 – present*
Accepted as incoming PhD student in the Machine Learning Department *2010 – future*
Previously enrolled in Masters in Language Technologies *2009 – 2010*
Stanford University, Stanford, CA: co-terminal master student *2001 – 2006*
Master of Science in Symbolic Systems
Bachelor of Science with honors in Symbolic Systems
Edina High School, Edina, MN *1998 – 2001*
University of Minnesota Talented Youth Mathematics Program, Minneapolis, MN *1997 – 2000*

Experience

Graduate Research Assistant, Carnegie Mellon Univ. Pittsburgh, PA *Sept 2009 – present*
Research assistant in statistical NLP group (Prof. Noah Smith), researching the prediction of financial indicators and public opinion through text regression. Completed submission to ICWSM on poll prediction from Twitter sentiment statistics, and further work on retail forecasting. Other projects include automated slang etymology mining via UrbanDictionary, noun phrase coreference resolution for information extraction, and general advising for several projects on Mechanical Turk annotation experiments.

Consulting Scientist, Dolores Labs (Crowdfower), Inc. San Francisco, CA *2008 – 2009*
Analysis, design, and engineering to use online workers from Amazon's Mechanical Turk service (an online labor marketplace) to solve customer annotation problems in information extraction, information retrieval, market research, and content moderation. Co-authored highly cited paper on applications to NLP. Wrote blog posts in the popular science style, attracting more than 200,000 views.

Software Engineer, Powerset, Inc. San Francisco, CA *2006 – 2008*
Designed and implemented many systems for a semantic search engine, especially relevance ranking and natural language engineering. Thorough end-to-end knowledge of highly complex question answering system. One of the first users of alpha-era Hadoop and EC2. Built practical data visualization and development tools that were still in use more than a year after I left, after the company's acquisition by Microsoft.

Research Assistant, Deme Project Stanford, CA *2003 – 2006*
Designed and developed an open source web application for online deliberation (discussion, document collaboration, voting).

Intern, Free University Berlin, Germany *2004*
Worked on e-Chalk educational instruction software, incorporating handwriting recognition.

Advising Fellow, Symbolic Systems Stanford, CA *2003 – 2005*
Undergraduate student advisor for Stanford's interdisciplinary Symbolic Systems Program. Academic peer advising, organizing and inviting speakers, running student and campus events, giving input to curriculum decisions.

Teaching, Stanford University Stanford, CA *2003 and 2005*
1) Section leader (i.e., TA) for introductory computer science course.
2) Developed and taught student-initiated reading course on *Gödel, Escher, Bach*, on the philosophy of artificial intelligence. Invited faculty speakers, led discussions.

Volunteer, Paul Wellstone for Senate Minneapolis, MN *July 2002*
Assisted in information management for U.S. Senate campaign.

Programming languages: R, Matlab, Prolog, C/C++, Java, Scala, Python, Ruby, Awk/Sed/Perl, PHP, Javascript, Flash/Flex.

Other technologies: MySQL, PostgreSQL, EC2, S3, Hadoop, Linux, LaTeX, Excel, Processing.

In Submission and Preparation

“From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.”

Brendan O’Connor, Ramnath Balasubramanyan, Bryan Routledge, Noah Smith.

Submitted to 4th International AAAI Conference on Weblogs and Social Media (ICWSM-2010).

“TweetMotif: Exploratory Search and Topic Summarization for Twitter.”

Brendan O’Connor, Michel Krieger, David Ahn.

Submitted to demo track of 4th International AAAI Conference on Weblogs and Social Media (ICWSM-2010).

Demo: <http://tweetmotif.com>

“The ARKref System for Noun Phrase Coreference Resolution.”

Brendan O’Connor and Michael Heilman. In preparation.

Demo: <http://www.ark.cs.cmu.edu/ARKref>

Refereed Conference Papers

“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. Empirical Methods in Natural Language

Processing (EMNLP), Honolulu, Hawaii, Oct 2008. Write-up and follow-ups at <http://blog.doloreslabs.com/?p=109>

“Groupware for Groups: Problem-Driven Design in Deme.”

Todd Davies, Brendan O’Connor, Alex Cochran, and Andrew Parker. Position paper from the Beyond Threaded Conversation Workshop at Computer-Human Interaction (CHI), Portland, Oregon, April 2005.

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, and Aaron Tam. In Todd Davies and Seeta Pena Gangadharan (editors), *Online Deliberation: Design, Research, and Practice*, CSLI Publications/University of Chicago Press, 2009.

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.

Invited Talks

“Dolores Labs and Amazon Mechanical Turk.”

Invited talk for Freebase (freebase.com), Metaweb Technologies, Inc., San Francisco, CA. May 2008.

“Human Computation for Defeating the Data Bottleneck: Dolores Labs and Amazon Mechanical Turk.”

Invited talk for the Stanford Natural Language Processing group (nlp.stanford.edu). May 2008.

“Fast, Cheap, and Good: Crowdsourcing with Amazon Mechanical Turk.”

Invited talk for the Stanford Human-Computer Interaction group (hci.stanford.edu). Sept 2008.

Tutorial

“An Interactive Introduction to R.”

Michael Driscoll and Brendan O'Connor.

Tutorial by Dataspora LLC, for Amyris, Inc., Emeryville, CA. June 2009. (dataspora.com, amyrisbiotech.com)

Academic Service

Program Committee for Workshop on Creating Speech and Language Data With Amazon's Mechanical Turk. The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics / Human Language Technologies, June 2010, Los Angeles.

Reviewer for Computer-Human Interaction (CHI-2010). 2009. (Submissions on Mechanical Turk.)

Awards

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.

Theses

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

Honors thesis, 2006. Readers: Noah Mark and James Fearon.

"Biased evidence assimilation under bounded Bayesian rationality."

M.S. thesis, 2006. Readers: Jonathan Bendor and James McClelland.

Other Projects

PalinSpeak.com. 2008.

Brendan O'Connor and Douglas Wilson.

Markov model chatbot as political satire.

Fleshmap.com. 2008.

With Fernanda Viegas and Martin Wattenberg.

Mechanical Turk data collection supporting the spatial visualization of human sexual preferences.

KDD Cup competition. 2009.

Task: Prediction system for marketing analysis.

Final rank: 287 of 4921 valid entries.

Netflix Prize competition. 2008.

Spent one month researching and implementing baseline recommendation system.

Final rank (2009): 645 out of 41305 teams.

Patents Pending

“Emphasizing Search Results According to Conceptual Meaning.”

Barney Pell, Scott Prevost, Giovanni Lorenzo Thione, Brendan O'Connor, Lukas Biewald.

Filed August 29, 2008. Application No. 12/201,504. Powerset, Inc. and Microsoft Corporation.

“Efficient Storage and Retrieval of Posting Lists.”

Chad Walters, Giovanni Lorenzo Thione, Barney Pell, Lukas Biewald, Brendan O’Connor.
Filed August 29, 2008. Application No. 12/201,079. 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.
Filed August 29, 2008. Application No. 12/201,978. Powerset, Inc. and Microsoft Corporation.

Selected Coursework

Carnegie Mellon (PhD)

Statistics and ML

Machine Learning (Carlos Guestrin)
Statistical Machine Learning (Larry Wasserman,
John Lafferty, in progress)

Statistical NLP

Language and Statistics II (Noah Smith, audited)
Advanced NLP Seminar (Noah Smith, audited)

Applications

Reading the Web (Tom Mitchell)
Text-Driven Forecasting (Noah Smith)
Social Media Analysis (William Cohen, in progress)
Graduate Research Assistant in Text-Based Financial
Forecasting (Noah Smith, Bryan Routledge)

Stanford (BS/MS)

Statistics and ML

Data Analysis (Jerome Friedman)
Statistical Inference (Joseph Romano)
Machine Learning (Andrew Ng)

Language and AI

Natural Language Processing (Christopher Manning)
Computational Linguistics (Martin Kay)
Introduction to Semantics and Pragmatics (David
Beaver)
Introduction to Syntax (Thomas Wasow, Ivan Sag)
Knowledge Representation and Reasoning (Tom
Costello)
Artificial Intelligence (Daphne Koller)

Rational Choice, Psychology, and Organizations

Economic Theory for Non-Economics PhD Students
Game Theory and Political Science (James Fearon)
New Economics of Organizations (Barry Weingast,
audited)
Behavioral Organization Theory (Jonathan Bendor)
Psychology and Economics (Antonio Rangel)
Neuroeconomics (Antonio Rangel, Brian Knutson)

Psychology, Sociology, Social Linguistics

Introduction to Social Networks (Sean Everton)
Cognitive Psychology (Todd Davies)
Language as a Public Concern (Geoffrey Nunberg)

Fundamental CS and Mathematics

Programming Paradigms
Object-Oriented Systems Design
Programming Languages
Discrete Mathematics
First-Order Logic
Introduction to Probability
Fundamental Concepts in Analysis
Multivariate Calculus (Univ. of Minnesota)
Linear Algebra (Univ. of Minnesota)
Cryptology and Number Theory (Univ. of
Minnesota)