

Dan Garrette

dhgarrette@gmail.com

<http://dhg.ai>

EDUCATION

<i>The University of Texas at Austin</i> , Austin, TX Ph.D., Computer Science Advisors: Jason Baldridge and Raymond Mooney	Aug 2009 - Apr 2015
<i>Illinois Wesleyan University</i> , Bloomington, IL B.S., Computer Science, <i>with Research Honors</i> Minor: Cognitive Science	Aug 2003 - Apr 2006 (completed in 3 years)

SELECTED RESEARCH EXPERIENCE

<i>Google Research</i> , New York, NY Senior Research Scientist Research Scientist · Machine learning and natural language processing.	Nov 2019 - Present Oct 2016 - Oct 2019
<i>University of Washington</i> , Seattle, WA Post-Doctoral Research Associate. Supervisor: Luke Zettlemoyer · Semi-supervised/low-resource learning for NLP.	May 2015 - Oct 2016
<i>University of Texas at Austin</i> , Austin, TX Research Assistant. Supervisor: Jason Baldridge · Learning NLP models from varieties of weak supervision.	Aug 2011 - May 2015
Research Assistant. Supervisors: Katrin Erk and Ray Mooney · Unifying logical and distributional semantics for natural language inference.	Aug 2009 - Aug 2011
<i>Google</i> , Mountain View, CA Intern · Machine learning and natural language processing research for Google News.	May 2013 - Aug 2013
<i>University of Maryland Institute for Advanced Computer Studies</i> , College Park, MD Research Assistant. Supervisor: Philip Resnik · Bayesian models of syntactic framing in political writing.	May 2012 - Aug 2012

SELECTED PUBLICATIONS

- [23] Rochelle Choenni, [Dan Garrette](#), Ekaterina Shutova. “How do languages influence each other? Studying cross-lingual data sharing during LLM fine-tuning”. In *Proc. of EMNLP*, 2023.
- [22] Rosanne Liu*, [Dan Garrette](#)*, Chitwan Saharia, William Chan, Adam Roberts, Sharan Narang, Irina Blok, RJ Mical, Mohammad Norouzi, Noah Constant*. “Character-Aware Models Improve Visual Text Rendering”. In *Proc. of ACL*, 2023.
- [21] Rochelle Choenni, [Dan Garrette](#), Ekaterina Shutova. “Cross-Lingual Transfer with Language-Specific Subnetworks for Low-Resource Dependency Parsing”. *Computational Linguistics*, 2023.
- [20] Parker Riley, Timothy Dozat, Jan A. Botha, Xavier Garcia, [Dan Garrette](#), Jason Riesa, Orhan Firat, Noah Constant. “FRMT: A Benchmark for Few-Shot Region-Aware Machine Translation”. *TACL*, 2023.
- [19] Jonathan H. Clark, [Dan Garrette](#), Iulia Turc, and John Wieting. “CANINE: Pre-training an Efficient Tokenization-Free Encoder for Language Representation”. *TACL*, 2022.
- [18] Jason Wei, [Dan Garrette](#), Tal Linzen, and Ellie Pavlick. “Frequency Effects on Syntactic Rule Learning in Transformers”. *EMNLP*, 2021.
- [17] Hyung Won Chung, [Dan Garrette](#), Kiat Chuan Tan, and Jason Riesa. “Improving Multilingual Models with Language-Clustered Vocabularies”. In *Proc. of EMNLP*, 2020.
- [16] Jonathan H. Clark, Eunsol Choi, Michael Collins, [Dan Garrette](#), Tom Kwiatkowski, Vitaly Nikolaev, and Jennimaria Palomaki. “TyDi QA: A Benchmark for Information-Seeking Question Answering in Typologically Diverse Languages”. *TACL*, 2020.
- [15] Telmo Pires, Eva Schlinger, and [Dan Garrette](#). “How multilingual is Multilingual BERT?” In *Proc. of ACL*, 2019.

- [14] Kelsey Ball and Dan Garrette. “Part-of-Speech Tagging for Code-Switched, Transliterated Texts without Explicit Language Identification”. In *Proc. of EMNLP*, 2018.
- [13] Maria Ryskina, Hannah Alpert-Abrams, Dan Garrette, and Taylor Berg-Kirkpatrick. “Automatic Composer Attribution in the First Folio of Shakespeare”. In *Proc. of ACL*, 2017.
- [12] Dan Garrette and Hannah Alpert-Abrams. “An Unsupervised Model of Orthographic Variation for Historical Document Transcription”. In *Proc. of NAACL*, 2016.
- [11] Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith. “A Supertag-Context Model for Weakly-Supervised CCG Parser Learning”. In *Proc. of CoNLL*, 2015.
- [10] Dan Garrette, Hannah Alpert-Abrams, Taylor Berg-Kirkpatrick, and Dan Klein. “Unsupervised Code-Switching for Multilingual Historical Document Transcription”. In *Proc. of NAACL*, 2015.
- [9] Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith. “Weakly-Supervised Grammar-Informed Bayesian CCG Parser Learning”. In *Proc. of AAAI*, 2015.
- [8] Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith. “Weakly-Supervised Bayesian Learning of a CCG Supertagger”. In *Proc. of CoNLL*, 2014.
- [7] Dan Garrette, Jason Mielens, and Jason Baldridge. “Real-World Semi-Supervised Learning of POS-Taggers for Low-Resource Languages”. In *Proc. of ACL*, 2013.
- [6] Dan Garrette and Jason Baldridge. “Learning a Part-of-Speech Tagger from Two Hours of Annotation”. In *Proc. of NAACL*, 2013.
 * **Best Talk Award** Finalist
- [5] Dan Garrette, Katrin Erk, and Raymond Mooney. “A Formal Approach to Linking Logical Form and Vector-Space Lexical Semantics”. Harry Bunt, Johan Bos, and Stephen Pulman (eds) *Computing Meaning, Vol. 4*, 2013.
- [4] Islam Beltagy, Cuong Chau, Gemma Boleda, Dan Garrette, Katrin Erk, and Raymond Mooney. “Montague Meets Markov: Deep Semantics with Probabilistic Logical Form”. In *Proc. of *SEM*, 2013.
- [3] Dan Garrette and Jason Baldridge. “Type-Supervised Hidden Markov Models for Part-of-Speech Tagging with Incomplete Tag Dictionaries”. In *Proc. of EMNLP*, 2012.
- [2] Dan Garrette, Katrin Erk, and Raymond Mooney. “Integrating Logical Representations with Probabilistic Information using Markov Logic”. In *Proc. of the Intl. Conference on Computational Semantics (IWCS)*, 2011.
- [1] Dan Garrette and Ewan Klein. “An Extensible Toolkit for Computational Semantics”. In *Proc. of the International Conference on Computational Semantics (IWCS)*, 2009.

INVITED TALKS

- *University of North Texas*. “Unsupervised Modeling for Historical Document Transcription”. Feb 2018.
- *NEH Reading the First Books Symp.*. “How to get a computer scientist involved in your DH project”. May 2017.
- *Ohio State University*. “Exploiting Universal Grammatical Properties to Induce CCGs”. March 2017.
- *Google Research*. “Exploiting Universal Grammatical Properties to Induce CCGs”. August 2016.
- *Amazon*. “Exploiting Universal Grammatical Properties to Induce CCGs”. August 2016.
- *Apple*. “Exploiting Universal Grammatical Properties to Induce CCGs”. August 2016.
- *Lawrence Livermore National Lab*. “Exploiting Universal Grammatical Properties to Induce CCGs”. August 2016.
- *Allen Institute for AI*. “Exploiting Universal Grammatical Properties to Induce CCGs”. August 2016.
- *University of Edinburgh*. “Learning CCGs from Weak Supervision”. June 2016.
- *Workshop on Multilingual and Crosslingual Methods in NLP (at NAACL-2016)*. “Unsupervised Modeling of Code-Switching and Orthographic Variation”. June 2016.
- *Microsoft Research*. “Unsupervised Modeling of Code-Switching and Orthographic Variation”. May 2016.
- *University of Washington*. “Learning CCGs from Weak Supervision”. February 2015.
- *Carnegie Mellon University*. “Learning CCGs from Weak Supervision”. April 2014.

ACADEMIC SERVICE

Conference reviewing: ACL, EMNLP, NAACL, EACL, ACL Rolling Review, COLING, and various workshops.
 Senior Area Chair: NAACL-2022 (Multilinguality).
 Area Chair: ACL, EMNLP, NAACL, COLING.
 Journal reviewing: TACL, Computational Linguistics, Language Resources and Evaluation.
 Conference website developer: IWCS-2013.
 UTCS Admissions Committee – Master’s Degree Program, Spring 2015.

SELECTED OPEN SOURCE PROJECTS

Numerous projects at <http://github.com/dhgarrette>

- Natural Language Toolkit (NLTK) – <http://www.nltk.org/> 2007 - 2010
- Toolkit for natural language processing in Python.
 - Authored most *semantics* code: first-order logic, λ -calculus, DRT, inference, etc.
- Ocular – <https://github.com/tberg12/ocular> 2014 - 2017
- State-of-the-art OCR system for transcribing historical texts.
 - Authored extensions/features used in our publications.
- Spring Batch – <http://projects.spring.io/spring-batch/> 2008 - 2009
- Application framework for batch processing in Java.

HONORS AND AWARDS*Graduate*

- Best Talk Award Finalist, NAACL-2013
- National Defense Science and Engineering Graduate Fellowship (NDSEG) - 2010–2013

Undergraduate

- Alumni Academic Scholarship
- Jennings Music Scholarship
- Upsilon Phi Epsilon - Computer Science Honor Society
- Dean's List - 6 of 6 semesters

OTHER SELECTED PROFESSIONAL EXPERIENCE*Accenture, Chicago, IL*

- Consultant Sep 2008 - Aug 2009
- Served as one of four committers to the open-source Spring Batch framework.
 - Designed and developed new functionality. Identified and fixed bugs.
 - Wrote reference documentation and answered questions on the public forum.
 - Assisted Accenture projects as a Subject Matter Expert for Spring Batch.
 - Designed Spring Batch training curriculum and led on-site training in the US and India.
 - Wrote documentation on Accenture's approach to Java development for use company-wide.
 - Certified by Accenture as a Technology Architect.
- Analyst May 2007 - Sep 2008
- Developed large-scale applications in Java.
 - After just four months, was made a sub-team leader, formally leading a group of developers.
 - After 15 months, promoted to Consultant. (Normal requirement is 2 to 3 years).

SELECTED TEACHING EXPERIENCE*The University of Texas at Austin*

- Instructor, *Natural Language Processing* Fall 2013
- Upper-division undergraduate Computer Science and Linguistics course.

The United States Peace Corps

- Volunteer, Ghana, West Africa 2006 - 2007
- Taught math and English in an underperforming rural junior secondary school.
 - Planned and taught HIV/AIDS presentations in rural communities and schools.

Illinois Wesleyan University

- Teaching Assistant 2004 - 2006
- Courses on algorithms (in C and Haskell), data structures (in C), and computer architecture.

ACTIVITIES

Post-Graduate

Seattle ScienceSlam - Presenter: *Learning to read 16th century books* (Voted **Best Talk** of the event) Jan 2016

Graduate

Organizer of the Natural Language Learning reading group 2011 - 2015

Undergraduate

John Wesley Powell Student Research Conference - Presenter

Symphonic Winds - Member, Percussion Section Leader for 5 semesters

Civic Orchestra, Opera Orchestra, Titan Band, Percussion Ensemble - Member