

## David McClosky

---

CONTACT INFORMATION	Google NYC 111 8th Avenue New York, NY 10011	Email: <a href="mailto:dmcc@cs.stanford.edu">dmcc@cs.stanford.edu</a> Website: <a href="http://nlp.stanford.edu/~mcclosky/">http://nlp.stanford.edu/~mcclosky/</a> Code: <a href="http://github.com/~dmcc/">http://github.com/~dmcc/</a>
RESEARCH INTERESTS	Information extraction, natural language parsing and syntax, question answering, semi-supervised methods, and domain adaptation.	
EDUCATION	<b>Brown University</b> Ph.D., Computer Science Thesis title: “Any Domain Parsing: Automatic Domain Adaptation for Natural Language Parsing” Committee: Eugene Charniak, Mark Johnson, and Dan Klein (University of California at Berkeley) Sc.M., Computer Science Thesis title: “Effective Self-training for Parsing” Committee: Eugene Charniak, Mark Johnson, and Tom Griffiths	Providence, RI, USA 9/2009 5/2006
	<b>University of Rochester</b> B.S. in Computer Science, <i>summa cum laude</i> , departmental honors Minors in linguistics and studio art.	Rochester, NY, USA 5/2004
RESEARCH EXPERIENCE	<b>Software engineer</b> , Google, New York, NY. <b>Research scientist</b> , DeepQA (Watson) project IBM Watson / Research, Yorktown, NY. <b>Postdoctoral scholar</b> , Natural Language Processing Group Department of Computer Science, Stanford University, Stanford, CA. <b>Visiting student</b> , Institute of Formal and Applied Linguistics (ÚFAL), Charles University, Prague, Czech Republic. <b>Graduate student</b> , Brown Laboratory for Linguistic Information Processing Department of Computer Science, Brown University, Providence, RI.	2/16–present 9/12–2/16 11/09–8/12 1/07–6/07 9/04–9/09
PUBLICATIONS	<b>Peer-reviewed journal articles</b> [1] David McClosky, Sebastian Riedel, Mihai Surdeanu, Andrew McCallum, and Christopher Manning. Combining joint models for biomedical event extraction. <i>BMC Bioinformatics</i> , 13(Suppl 11):S9, 2012. <b>Peer-reviewed conference articles</b> [2] Oren Melamud, David McClosky, Siddharth Patwardhan, and Mohit Bansal. The role of context types and dimensionality in learning word embeddings. In <i>Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies</i> , pages 1030–1040. Association for Computational Linguistics, 2016.	

- [3] Do Kook Choe, David McClosky, and Eugene Charniak. Syntactic parse fusion. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP 2015)*, Lisbon, Portugal, September 2015.
- [4] Do Kook Choe and David McClosky. Parsing paraphrases with joint inference. In *Proceedings of 53rd Annual Meeting of the Association for Computational Linguistics (ACL 2015)*, Beijing, China, July 2015.
- [5] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60, Baltimore, Maryland, June 2014.
- [6] David McClosky and Christopher D. Manning. Learning constraints for consistent timeline extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Natural Language Learning (EMNLP-CoNLL 2012)*, Jeju, Korea, July 2012.
- [7] David McClosky, Mihai Surdeanu, and Christopher D. Manning. Event extraction as dependency parsing. In *Proceedings of the Association for Computational Linguistics - Human Language Technologies 2011 Conference (ACL-HLT 2011), Main Conference*, Portland, Oregon, June 2011.
- [8] David McClosky, Mihai Surdeanu, and Christopher D. Manning. Event Extraction as Dependency Parsing for BioNLP 2011. In *Proceedings of the BioNLP Workshop*, June 2011.
- [9] Sebastian Riedel, David McClosky, Mihai Surdeanu, Andrew McCallum, and Christopher D. Manning. Model Combination for Event Extraction in BioNLP 2011. In *Proceedings of the BioNLP Workshop*, June 2011.
- [10] Mihai Surdeanu, David McClosky, Mason R. Smith, Andrey Gusev, and Christopher D. Manning. Customizing an information extraction system to a new domain. In *Proceedings of the Workshop on Relational Models of Semantics*, June 2011.
- [11] David McClosky, Eugene Charniak, and Mark Johnson. Automatic domain adaptation for parsing. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies 2010 Conference (NAACL-HLT 2010), Main Conference*, Los Angeles, California, June 2010.
- [12] William P. Headden III, Mark Johnson, and David McClosky. Improving unsupervised dependency parsing with richer contexts and smoothing. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2009), Main Conference*, Boulder, Colorado, May 2009.
- [13] David McClosky, Eugene Charniak, and Mark Johnson. When is self-training effective for parsing? In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Manchester, UK, August 2008.
- [14] William P. Headden III, David McClosky, and Eugene Charniak. Evaluating unsupervised part-of-speech tagging for grammar induction. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, Manchester, UK, August 2008.
- [15] David McClosky and Eugene Charniak. Self-training for biomedical parsing. In *Proceedings of the Human Language Technology Conference of the Association for Computational Linguistics (HLT-ACL 2008), Short Papers*, pages 101–104, Columbus, Ohio, June 2008. Association for Computational Linguistics.

- [16] Matthew Lease, Eugene Charniak, Mark Johnson, and David McClosky. A look at parsing and its applications. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence (AAAI 2006)*, July 2006.
- [17] David McClosky, Eugene Charniak, and Mark Johnson. Reranking and self-training for parser adaptation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING-ACL 2006)*, pages 337–344, Sydney, Australia, July 2006. Association for Computational Linguistics.
- [18] David McClosky, Eugene Charniak, and Mark Johnson. Effective self-training for parsing. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2006), Main Conference*, pages 152–159, New York City, USA, June 2006. Association for Computational Linguistics.
- [19] Sharon Goldwater and David McClosky. Improving statistical MT through morphological analysis. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP 2005)*, pages 676–683, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

#### Technical reports and other publications

- [20] David McClosky, Wanxiang Che, Marta Recasens, Mengqiu Wang, Richard Socher, and Christopher D. Manning. Stanford’s System for Parsing the English Web. Notes of the First Workshop on Syntactic Analysis of Non-Canonical Language (SANCL), June 2012.
- [21] Mihai Surdeanu, Sonal Gupta, John Bauer, David McClosky, Angel X. Chang Valentin I. Spitzkovsky, and Christopher D. Manning. Stanford’s distantly-supervised slot-filling system. In *Proceedings of the Knowledge Base Population Track of the Text Analysis Workshop (TAC-KBP’11)*, Gaithersburg, Maryland, November 2011.
- [22] Mihai Surdeanu, David McClosky, Julie Tibshirani, John Bauer, Angel Chang, Valentin I. Spitzkovsky, and Christopher D. Manning. A simple distant supervision approach for the kbp slot filling task. In *Proceedings of the Knowledge Base Population Track of the Text Analysis Workshop (TAC-KBP’10)*, Gaithersburg, Maryland, November 2010.
- [23] David McClosky. Modeling valence effects in unsupervised grammar induction. Technical Report CS-09-01, Brown University, Providence, RI, USA, 2008.

#### Datasets

- [24] David McClosky, Eugene Charniak, and Mark Johnson. *BLLIP North American News Text*, 2008. Linguistic Data Consortium. LDC2008T13, LDC2008T14.

#### INVITED TALKS AND PANELS

- “The Role of Context Types and Dimensionality in Learning Word Embeddings,”  
Text IQ, New York, NY. 7/29/2016
- “Field to Forest: Parsing in Unfamiliar Territory,”  
Google, New York, NY. 11/20/2015

“Field to Forest: Parsing in Unfamiliar Territory,” Columbia University, New York, NY.	11/16/2015
“Field to Forest: Parsing in Unfamiliar Territory,” Amazon, Sunnyvale, CA.	11/10/2015
“Field to Forest: Parsing in Unfamiliar Territory,” Yahoo!, Sunnyvale, CA.	11/6/2015
“Field to Forest: Parsing in Unfamiliar Territory,” IBM Research, Dublin, Ireland.	9/24/2015
“Field to Forest: Parsing in Unfamiliar Territory,” Dublin City University, Dublin, Ireland.	9/23/2015
“Co-dependent Information Extraction,” CUNY Grad Center, New York, NY.	4/26/2013
“Co-dependent Information Extraction,” Columbia University, New York, NY.	4/15/2013
“Parsing to Extract Meaning,” Microsoft Research, Redmond, WA.	4/16/2012
“Parsing to Extract Meaning,” Google Research, Mountain View, CA.	4/13/2012
“Parsing to Extract Meaning,” IBM Research, Hawthorne, NY.	3/26/2012
Tutorial on Stanford CoreNLP and TAC-KBP, University of Wisconsin Madison, Madison, WI.	4/22/2011
“Event Extraction as Dependency Parsing,” University of Wisconsin Madison, Madison, WI.	4/21/2011
“Event Extraction as Dependency Parsing,” University of Massachusetts Amherst, Amherst, MA.	2/24/2011
“Automatic Domain Adaptation for Parsing,” University of California at Berkeley, Berkeley, CA.	9/28/2009
Semi-supervised Learning for Natural Language Processing Workshop, Panelist, Boulder, CO.	6/4/2009
OTHER TALKS AND PRESENTATIONS	
“How to Understand Biomedical Text (And More),” Stanford University, Stanford, CA.	5/1/2012
“Joint inference for Consistent Timeline Extraction,” Stanford University, Stanford, CA.	12/8/2011
“Event Extraction as Dependency Parsing,” Poster at DARPA Machine Reading Phase 3 Kickoff, Seattle, WA.	4/6/2011

	“Dependency Parsing for Hierarchical Event Extraction,” Stanford University, Stanford, CA.	10/28/2010
	“Fast, Accurate Parsing,” Poster at DARPA Machine Reading Phase 2 Kickoff, St. Petersburg, FL.	4/15/2010
	“Automatic Domain Adaptation for Parsing,” Stanford University, Stanford, CA.	1/21/2010
	“Towards Any Domain Parsing,” Poster at Brown-MIT-Harvard NLP Summit, MIT, Cambridge, MA.	5/28/2009
	“Hogwash: Simple parallel programming for everyone,” Parallel computing package tutorial, Brown University, Providence, RI.	2/10/2009
	“Valency Effects in Grammar Induction,” Annual International Research Training Group (IRTG) meeting, Language Technology and Cognitive Systems, Saarland University and University of Edinburgh, Trier, Germany.	6/6/2007
	“Valency Effects in Grammar Induction,” Institute of Formal and Applied Linguistics (ÚFAL), Charles University, Prague, Czech Republic.	4/26/2007
	“Proposed Extensions for Dependency Grammar Induction,” Poster at Brown-MIT-Harvard NLP Summit, MIT, Cambridge, MA.	11/10/2006
HONORS AND AWARDS	<b>van Dam Fellowship</b> , Brown University	2009
	<b>University Fellowship</b> , Brown University (1 year)	2004
	<b>Phi Beta Kappa</b>	2003
	<b>Golden Key</b>	2003
MENTORING	Oren Melamud, Bar-Ilan University	Summer 2015
	Do Kook Choe, Brown University	Summer 2014
PROFESSIONAL SERVICE	<b>Conference area chair:</b> EACL (Information Extraction)	2012, 2014
	<b>Workshop co-organizer:</b> ACL Domain Adaptation for Natural Language Processing (DANLP)	2010
	<b>Grant reviewing:</b> Israeli Science Foundation	2011, 2014
	<b>Session chair:</b> EMNLP	2015
	<b>Senior program committee:</b> IJCAI	2011
	<b>Conference, journal, and workshop reviewing:</b> AAI (2008, 2013), ACL (2009–2016), ACL Semi-supervised Learning Workshop (2009), ACL Student Research Workshop (2009, 2011), ACM TIST (2010), AIRS (2012), BioNLP shared task (2011, 2013, 2016), COLING (2008–2016), CoNLL (2011–2016), EACL (2009), EMNLP (2007–2008, 2010–2016), ESSLI workshop on extrinsic parse improvement (2013), IJCNLP (2011), IWPT (2011, 2015), JAIR (2009), JMLR (2010), Journal of Biomedical Semantics (2015),	

LRE (2007), NAACL (2009–2010, 2013, 2015–2016), NAACL Student Research Workshop (2009), Semantic Mining in Biomedicine (2012, 2014), SPRML-SANCL (2014), SPRML (2015), WWW (2015–2016)

**Organizer:** Brown Machine Learning Reading Group 1/08–1/09  
IBM DeepQA Reading Group 5/13–2/16

**Student volunteer:** HLT-EMNLP (2005), NAACL-HLT (2009)

**LaTeX consultant:** Brown University, Computer Science Department 11/05–12/06

**President:** Computer Science Undergraduate Council, University of Rochester 2004

#### SOFTWARE AND MODELS

*BLLIP (Charniak and Johnson) reranking parser* (maintenance and extensions)  
<http://github.com/BLLIP/bllip-parser/>

*Stanford CoreNLP* (contributor)  
<http://nlp.stanford.edu/software/corenlp.shtml>

*Natural Language Toolkit (NLTK)* (contributor)  
<http://nltk.org/>

*Stanford Biomedical Event Parser* (co-author)  
<http://nlp.stanford.edu/software/eventparser.shtml>

*PyStanfordDependencies*  
<http://github.com/dmcc/PyStanfordDependencies/>

*Ensemble: Linearly-interpolated Dependency Parsers* (contributor)  
<http://www.surdeanu.info/mihai/ensemble/>

*Newswire self-trained parsing model*  
<http://stanford.edu/~mcclosky/selftraining/>

*Biomedical self-trained parsing model*  
<http://stanford.edu/~mcclosky/biomedical/>

*parsedyff*: Parse tree difference visualizer  
<http://stanford.edu/~mcclosky/software/parsedyff/>

*Hogwash*: Parallel computing package for Python  
<http://cs.brown.edu/~dmcc/hogwash/>