

David McClosky

- CONTACT INFORMATION** Department of Computer Science Email: dmcc@cs.brown.edu
Brown University Website: <http://www.cs.brown.edu/~dmcc/>
115 Waterman Ave, 4th Floor Fax: +1 (401) 863-7657
Providence, RI 02912-1910
- RESEARCH INTERESTS** Semi-supervised and unsupervised learning of natural language, integration of phonology, morphology, and syntax
- EDUCATION**
- Brown University** Providence, RI, USA
Ph.D., Computer Science (expected) September 2009
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 May 2006
Thesis title: “Effective Self-training for Parsing”
Committee: Eugene Charniak, Mark Johnson, and Tom Griffiths
- University of Rochester** Rochester, NY, USA
B.S. in Computer Science, summa cum laude, departmental honors May 2004
Minors in linguistics and studio art.
- RESEARCH EXPERIENCE**
- Graduate student**, Brown Laboratory for Linguistic Information Processing (BLLIP)
Department of Computer Science, Brown University, Providence, RI. 9/04–Present
- Visiting student**, Institute of Formal and Applied Linguistics (ÚFAL), Charles University, Prague, Czech Republic. 1/07–6/07
- PUBLICATIONS**
- Conference Proceedings**
- [1] 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'09)*, Main Conference, Boulder, Colorado, May 2009.
 - [2] 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'08)*, Manchester, UK, August 2008.
 - [3] 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'08)*, Manchester, UK, August 2008.
 - [4] David McClosky and Eugene Charniak. Self-training for biomedical parsing. In *Proceedings of the Human Language Technology Conference of the the Association for Computational Linguistics (HLT-ACL'08)*, Short Papers, pages 101–104, Columbus, Ohio, June 2008. Association for Computational Linguistics.

- [5] 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'06)*, 16–20 July 2006.
- [6] 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'06)*, pages 337–344, Sydney, Australia, July 2006. Association for Computational Linguistics.
- [7] 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'06), Main Conference*, pages 152–159, New York City, USA, June 2006. Association for Computational Linguistics.
- [8] 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'05)*, pages 676–683, Vancouver, British Columbia, Canada, October 2005. Association for Computational Linguistics.

Technical Reports

- [9] David McClosky. Modeling valence effects in unsupervised grammar induction. Technical Report CS-09-01, Brown University, Providence, RI, USA, 2008.

Data Sets

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

INVITED TALKS AND PANELS	Semi-supervised Learning for Natural Language Processing Workshop, Panelist, Boulder, CO	6/4/2009
OTHER TALKS AND PRESENTATIONS	“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	van Dam Fellowship , Brown University University Fellowship , Brown University (1 year) Phi Beta Kappa Golden Key	2009 2004 2003 2003

PROFESSIONAL
SERVICE

Organizer: Brown Machine Learning Reading Group 1/08–1/09

Program committee: ACL Semi-supervised Learning Workshop, ACL Student Research Workshop, NAACL Student Research Workshop 2009

Reviewer: AAAI (2008), ACL (2009), COLING (2008), EACL (2009), EMNLP (2007–2008), JAIR (2009), LRE (2007), NAACL (2009) 2007–Present

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

Newsire self-trained parsing model
<http://cs.brown.edu/~dmcc/selftraining/>

Biomedical self-trained parsing model
<http://cs.brown.edu/~dmcc/biomedical/>

parsedyff: Parse tree difference visualizer
<http://cs.brown.edu/~dmcc/software/parsedyff/>

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