Matthew G. Snover

University of Maryland, College Park
Department of Computer Science
A.V. Williams Building
College Park, MD 20742
Office: (203) 691-6113
Email: snover@cs.umd.edu

95 Carmalt Rd
Hamden, CT 06517
Date of Birth: April 16, 1978.
Citizenship: U.S.A.
Homepage: www.cs.umd.edu/~snover

Research Interests

Computational Linguistics, Statistical Machine Translation, Machine Learning, Unsupervised and Minimally Supervised Learning, Evaluation Metrics for Natural Language Processing, and Artificial Intelligence

Education

Ph.D. Computer Science, University of Maryland, College Park, *expected May 2010*.

Dissertation: Improving Machine Translation Using Comparable Corpora

Advisor: Professor Bonnie Dorr

M.S. Computer Science, Washington University in St. Louis, 2002.

Master's Thesis: An Unsupervised Knowledge Free Algorithm for the Learning of Morphology in Natural Languages

Advisor: Professor Michael R. Brent

B.S. Computer Science, Washington University in St. Louis, 2000.

Second Major: Philosophy

Minor: Psychology

Employment

Graduate Research Assistant 2003–Present. University of Maryland, College Park.

Member of BBN Technologies' AGILE MT Initiative, GALE Research Program. 2005–Present. Worked as an integral part of the core Machine Translation team at BBN Technologies, to bring my research to fruition in a real application as part of the top-performing MT system in the GALE program. Also led effort to prototype and develop the official TER and HTER metrics used by the GALE program, and spearheaded UMD/BBN's top-performing entry (TERp) in the NIST Metrics MATR 2008 challenge.

Member of BBN Technologies' Blue Team, EARS Research Program. 2003-2005.

Graduate Teaching Assistant Fall 2002. University of Maryland, College Park.

Graduate Research Assistant 2000–2002. Washington University in St Louis.

Grader 1997–2000. Washington University in St Louis.

Lab Assistant Intern 1995-1996. Genetics Institute, Cambridge MA.

Duties primarily consisted of programming robotic systems for automated assays, running robotic assays and data analysis of results.

Teaching

Tutorial Co-instructor *(Summer 2005)*

Johns Hopkins University Center for Language and Speech Processing (CLSP) Summer School

Graduate Teaching Assistant *(Fall 2002)*

Department of Computer Science, University of Maryland, College Park.

Teaching assistant for CMSC 330 "Organization of Programming Languages" *Fall 2002*. Teaching duties included discussion sections covering instruction in programming Perl, Lisp, and Prolog.

Grader *(1997 – 2000)*

Department of Computer Science, Washington University in St. Louis.

Grader for CS 201: Formal Foundations of Computer Science (2 semesters), CS 342: Object-Oriented Software Development Laboratory, CS 455: Programming Systems and Languages (4 semesters), and CS 504: Programming Concepts and Practice.

Selected Publications

Statistical Machine Translation and Machine Translation Evaluation

To Appear 2010 **Matthew Snover**, Nitin Madnani, Bonnie Dorr and Richard Schwartz. "TER-Plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate". *Machine Translation Journal, Special Issue on: Automated Metrics for MT Evaluation.*

To Appear 2010 Jeremy G Kahn, **Matthew Snover**, and Mari Ostendorf. "Expected Dependency Pair Match: Predicting translation quality with expected syntactic structure." *Machine Translation Journal, Special Issue on: Automated Metrics for MT Evaluation.*

2009 **Matthew Snover**, Nitin Madnani, Bonnie J. Dorr, and Richard Schwartz. "Fluency, Adequacy, or HTER? Exploring Different Human Judgments with a Tunable MT Metric." Proceedings of the Fourth Workshop on Statistical Machine Translation at the 12th Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2009), 2009.

2008 **Matthew Snover**, Bonnie J. Dorr, and Richard Schwartz. "Language and Translation Model Adaptation using Comparable Corpora." Proceedings of the Conference on Empirical Methods in Natural Language Processing 2008 (EMNLP 2008), 2008.

2006 **Matthew Snover**, Bonnie J. Dorr, Richard Schwartz, Linnea Micciulla, and John Makhoul. "A Study of Translation Edit Rate with Targeted Human Annotation." Proceedings of Association for Machine Translation in the Americas, 2006.

Parsing and Metadata in Automatic Speech Recognition

2006 John Hale, Izhak Shafran, Lisa Yung, Bonnie Dorr, Mary Harper, Anna Krasnyanskaya, Matthew Lease, Yang Liu, Brian Roark, **Matthew Snover**, and Robin Stewart. "PCFGs with Syntactic and Prosodic Indicators of Speech Repairs." Proceedings of the Association of Computational Linguistics 2006 (ACL 2006), 2006.

2006 Brian Roark, Yang Liu, Mary Harper, Robin Stewart, Matthew Lease, **Matthew Snover**, Izhak Shafran, Bonnie J. Dorr, John Hale, Anna Krasnyanskaya, and Lisa Yung. "Reranking for Sentence Boundary Detection in Conversational Speech." Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2006), Toulouse France, 2006.

2006 Brian Roark, Mary Harper, Eugene Charniak, Bonnie J. Dorr, Mark Johnson, Jeremy G. Kahn, Yang Liu, Mari Ostendorf, John Hale, Anna Krasnyanskaya, Matthew Lease, Izhak Shafran, **Matthew Snover**, Robin Stewart, and Lisa Yung. "Sparseval: Evaluation Metrics for Parsing Speech." Proceedings of the International conference on Language Resources and Evaluation (LREC 2006), Genoa, Italy, 2006.

2005 Mary Harper, Bonnie J. Dorr, John Hale, Brian Roark, Izhak Shafran, Matthew Lease, Yang Liu, **Matthew Snover**, Lisa Yung, Anna Krasnyanskaya, and Robin Stewart. "Final Report on Parsing and Spoken Structural Event Detection", 2005 Johns Hopkins Summer Workshop, 2005.

2004 **Matthew Snover**, Richard Schwartz, Bonnie Dorr, and John Makhoul. "RT-S: Surface Rich Transcription Scoring, Methodology, and Initial Results." Proceedings of the Rich Transcription 2004 Workshop.

2004 **Matthew Snover**, Bonnie Dorr, and Richard Schwartz. "A Lexically-Driven Algorithm for Disfluency Detection." Short Paper Proceedings of North American Association for Computational Linguistics (NAACL) and Human Language Technology (HLT) Conference 2004

Unsupervised Learning of Morphology

2002 **Matthew Snover**. "An Unsupervised Knowledge Free Algorithm for the Learning of Morphology in Natural Languages." Washington University in St Louis, Department of Computer Science, MS Thesis. 2002.

2002 **Matthew Snover** and Michael Brent. "A Probabilistic Model for Learning Concatenative Morphology." Proceedings of Neural Information Processing Systems (NIPS) 2002.

2002 **Matthew Snover**, Gaja Jarosz, and Michael Brent. "Unsupervised Learning of Morphology Using a Novel Directed Search Algorithm: Taking the First Step." Association for Computational Linguistics (ACL-2002): Workshop on Morphological and Phonological Learning. 2002.

2001 **Matthew Snover** and Michael Brent. "A Bayesian Model for Morpheme and Paradigm Identification." Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL-2001), pages 482-490. 2001.

In Preparation

Machine Translation Evaluation and Optimization. *(Book chapter in preparation)*.

Targeted Translation Model Adaptation: Learning Translation Rules from Monolingual Text. *(Journal article in preparation)*

Public Software Releases

Lead Developer - Translation Edit Rate plus (*TERp*)

Software to evaluate the quality of Machine Translation output using numerous enhancements to the Translation Edit Rate framework, including synonymy, stemming, paraphrases, and edit cost optimization. TERp was the highest overall performing evaluation metric at the NIST Metrics MATR 2008 Challenge.

Software Webpage: `www.umiacs.umd.edu/~snover/terp/`

Copyright: BBN Technologies and University of Maryland

Initial Release: March 2009

Lead Developer - Translation Edit Rate Computation (*TERcom*)

Software to evaluate the quality of Machine Translation output according to the Translation Edit Rate evaluation metric. This software is the official scoring tool for DARPA's Global Automatous Language Exploitation program (GALE), and has been incorporated into NIST's Human Targeted Reference Annotation tool. Since its release, TER has been become one of most of the commonly used evaluation metrics for Machine Translation, and has been used as a general purpose alignment tool for many research projects outside of Machine Translation Evaluation.

Software Webpage: `www.cs.umd.edu/~snover/tercom/`

Copyright: BBN Technologies and University of Maryland

Initial Release: 11/10/05 (Perl version), 4/28/06 (Java version)

Principal Service Activities

Graduate Student Participant: Johns Hopkins University Center for Language and Speech Processing (CLSP) Summer Workshop 2005, "Parsing and Spoken Structural Event Detection"

Panelist: GALE Program PI Meeting, 2008, Panel on "Automatic Machine Translation Evaluation"

Program Committee Member: Conference on Empirical Methods in Natural Language Processing (EMNLP), 2009

Referee: Journal of Machine Translation

Miscellaneous

Computer Skills: C/C++, Java, Perl, Gawk, LaTeX, Scheme, Lisp.

Last updated: January 19, 2010

`www.cs.umd.edu/~snover/snover_cv.pdf`