

## REBECCA HWA

Department of Computer Science  
Sennott Square Building, Room 5421  
210 S. Bouquet St  
University of Pittsburgh  
Pittsburgh, PA 15260, U.S.A.

*phone:* +1 (412) 624-7953  
*fax:* +1 (412) 624-8854  
*email:* hwa@cs.pitt.edu  
*WWW:* <http://www.cs.pitt.edu/~hwa/>

**Education** HARVARD UNIVERSITY Ph.D. March 2001  
Division of Engineering and Applied Sciences, field of Computer Science.  
Thesis: *Learning Probabilistic Lexicalized Grammars for Natural Language Processing*

HARVARD UNIVERSITY S.M. May 1994  
Division of Engineering and Applied Sciences, field of Computer Science.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES B.S. June 1993  
Summa cum laude in Computer Science and Engineering.

**Employment** UNIV. OF PITTSBURGH  
DEPARTMENT OF COMPUTER SCIENCE Fall 2003 – present  
Pittsburgh, PA  
Assistant professor.

UNIV. OF MARYLAND  
INSTITUTE FOR ADVANCED COMPUTER STUDIES Winter 2000 – Spring 2003  
College Park, MD  
Postdoctoral research associate. Developed learning algorithms for cross-language processing and machine translation.

DIGITAL EQUIPMENT CORPORATION CAMBRIDGE RESEARCH LAB Summer 1997  
Cambridge, MA  
Intern. Designed a hierarchical physical model for human facial animation. The system controls low-level motor skills, freeing the animators from tedious direct manipulations.

APPLIED LANGUAGE TECHNOLOGIES 1995 – 1996  
Cambridge, MA  
Independent Contractor. Conducted experiments to test out different noise cancellation techniques to enhance clarity over telephone lines for better speech recognition performance.

TICA TECHNOLOGIES, INC. Summer 1994  
Cambridge, MA  
Independent Contractor. Developed an algorithm that organized hypermedia documents to facilitate the efficiency of user navigations.

MICROSOFT CORPORATION  
Redmond, WA

Summer 1993

Software Development Intern. Implemented an anti-aliasing algorithm and optimized it for efficiency to facilitate real-time pen inking of anti-aliased line drawings under the Pen Windows platform.

CORNELL UNIVERSITY  
Ithaca, NY

Summer 1992

Supercomputing Program for Undergraduate Research. Built computational models of the human heart to simulate its electrical properties in three dimensional space. Visualized the results of the simulation with the Data Explorer visualization package.

MICROSOFT CORPORATION  
Redmond, WA

Summer 1991

Software Development Intern. Optimized the spelling checker for Word for the Macintosh for speed; implemented the control resources for ResEdit that interface between Word and its Installable Commands.

## Journal Articles

K. Liu, W. Chapman, R. Hwa, and R. Crowley. 2007. Heuristic Sample Selection to Minimize Reference Standard Training Set for a Part-of-Speech Tagger. To appear in *Journal of the American Medical Informatics Association*.

T. Wilson, J. Wiebe, and R. Hwa. 2006. Recognizing Strong and Weak Opinion Clauses. In *Computational Intelligence*, 22:2, pp.73 –99.

R. Hwa, P. Resnik, A. Weinberg, C. Cabezas, and O. Kolak. 2005. Bootstrapping Parsers via Syntactic Projection across Parallel Texts. In *Journal of Natural Language Engineering*, special issue on parallel text. Eds. Rada Mihalcea and Michel Simard. 11:3, pp. 311-325.

R. Hwa. Sample Selection for Statistical Parsing. 2004. In *Computational Linguistics*, 30:3.

B. Dorr, N.F. Ayan, N. Habash, N. Madnani, and R. Hwa, Rapid Porting of DUSTER to Hindi. 2003. In *ACM Transactions on Asian Language Information Processing (TALIP)*, 2:3.

N.F. Otani and R. Hwa. 1994. Topological Considerations in the Generation of Scroll Waves in Excitable and Cyclical Media. *Physica D*, 77, pp. 434-455.

## Refereed Conference Papers

J. Albrecht and R. Hwa. 2007. A Re-examination of Machine Learning Approaches for Sentence-Level MT Evaluation. In the Proceedings of the 45th Annual Meeting of the ACL, Prague, Czech Republic.

J. Albrecht and R. Hwa. 2007. Regression for Sentence-Level MT Evaluation with Pseudo References. In the Proceedings of the 45th Annual Meeting of the ACL, Prague, Czech Republic.

M. Mylonakis, K. Sima'an, and R. Hwa. 2007. Unsupervised Estimation for Noisy-Channel Models. To appear in Proceedings of International Conference of Machine Learning (ICML), Corvallis, Oregon.

R. Hwa, C. Nichols, K. Sima'an. 2006. Corpus Variations for Translation Lexicon Induction. In the Proceedings of the seventh conference of the Association for Machine Translation in the Americas. Cambridge, MA.

C. Xi and R. Hwa. 2005. A Backoff Model for Bootstrapping Resources for Non-English Languages. In the Proceedings of HLT/EMNLP-05, Vancouver, Canada.

T. Wilson, J. Wiebe, and R. Hwa. 2004. Just how mad are you? Finding strong and weak opinion clauses. In the Proceedings of AAAI-2004, San Jose, California.

M. Steedman, R. Hwa, S. Clark, M. Osborne, A. Sarkar, J. Hockenmaier, P. Ruhlén, S. Baker, and J. Crim. 2003. Example Selection for Bootstrapping Statistical Parsers. In the Proceedings of the Joint Conference of Human Language Technologies and the Annual Meeting of the North American Chapter of the ACL, Edmonton, Canada.

M. Steedman, M. Osborne, A. Sarkar, S. Clark, R. Hwa, J. Hockenmaier, P. Ruhlén, S. Baker, and J. Crim. 2003. Bootstrapping Statistical Parsers from Small Datasets. In the Proceedings of the Tenth Conference of the European Chapter of the ACL, Budapest, Hungary.

R. Hwa, P. Resnik, A. Weinberg, and O. Kolak. 2002. Evaluating Translational Correspondence using Annotation Projection. In the Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania.

B. Dorr, L. Pearl, R. Hwa, and N. Habash. 2002. DUSTER: A Method for Unraveling Cross-Language Divergences for Statistical Word-level Alignment. In the Proceedings of the Fifth Conference of the Association for Machine Translation in the Americas (AMTA-2002), Tiburon, California.

R. Hwa. 2001. On Minimizing Training Data for Parser Acquisition. In the Proceedings of the Conference on Computational Natural Language Learning (CoNLL-2001).

R. Hwa. 2000. Sample Selection for Grammar Induction. In the Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, pp. 45-52, Hong Kong, China.

R. Hwa. 1999. Supervised Grammar Induction using Training Data with Limited Constituent Information. In the Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, pp. 73-77, College Park, Maryland.

R. Hwa. 1998. An Empirical Evaluation of Probabilistic Lexicalized Tree Insertion Grammars. In the Proceedings of COLING-ACL 1998, volume 1, pp. 557-563, Montreal, Canada.

**Refereed  
Workshop  
Papers**

B. Mohit and R. Hwa. 2007. Localization of Difficult-to-Translate Phrases. In the Proceedings of the 2nd ACL Workshop on Statistical Machine Translations, Prague, Czech Republic.

B. Mohit and R. Hwa. 2005. Syntax-based Semi-Supervised Named Entity Tagging. In the Proceedings of the 43rd Meeting of the ACL (short paper track companion volume), Ann Arbor, Michigan.

C. Nichols and R. Hwa. 2005. Word Alignment and Cross-Lingual Resource Acquisition. In the Proceedings of the 43rd Meeting of the ACL (short paper track companion volume), Ann Arbor, Michigan.

B. Maceireizo, D. Litman, and R. Hwa. 2004. Co-training for Predicting Emotions with Spoken Dialogue Data. In the Companion Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (Student Research Workshop).

R. Hwa, M. Osborne, A. Sarkar, and M. Steedman. 2003. Corrected Co-training for Statistical Parsers. In the Proceedings of the Workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining, International Conference of Machine Learning, Washington D.C..

R. Hwa, P. Resnik, and A. Weinberg. 2002. Breaking the Resource Bottleneck for Multilingual Parsing. In the Proceedings of the Workshop on *Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data*, Third International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Canary Islands, Spain.

A. Lopez, M. Nossal, R. Hwa, and P. Resnik. 2002. Word-level Alignment for Multilingual Resource Acquisition. In the Proceedings of the Workshop on *Linguistic Knowledge Acquisition and Representation: Bootstrapping Annotated Language Data*, Third International Conference on Language Resources and Evaluation (LREC-2002), Las Palmas, Canary Islands, Spain.

**Invited  
Talks**

“Semi-supervised Learning for Information Management”

University of Maryland, College Park. February, 2005.

“Breaking the Resource Bottleneck for Multilingual Processing,”

University of Edinburgh, UK, IGK Summer School. September, 2004.

“A Continuum of Bootstrapping Methods for Statistical Parsing,”

University of Rochester, NY, December 2003.

“Semi-Supervised Approaches for Learning to Parse Natural Languages,”

University of Michigan, MI, April 2003.

Information Sciences Institute, CA, April 2003.

University of Pittsburgh, PA, March 2003.

“Sample Selection for Parser Induction,”

University of Pennsylvania, Philadelphia, PA, January 2002.

“Learning Probabilistic Lexicalized Grammars for Natural Language Processing,”

Johns Hopkins University, Baltimore, MD, March 2001.

Naval Research Laboratory, Washington D.C., July 2000.

IBM Watson Research Laboratory, Yorktown Heights, NY, June 2000.

University of Maryland, College Park, MD, June 2000.

University of Chicago, IL, April 2000.

**Patent**

Apparatus for determining the structure of a hypermedia document using graph partitioning (with Joe Marks and Stuart Shieber). U. S. patent number 5,546,517, August 1996.

**Teaching Experience** UNIV. OF PITTSBURGH Fall 2003 - present  
Pittsburgh, PA  
Courses taught include: a graduate level course on machine learning, a graduate level seminar on natural language processing, an upper division undergraduate course on artificial intelligence, an upper division undergraduate course on computer graphics, and a required undergraduate writing course on the social implications of computing.

UNIV. OF MARYLAND Spring 2002  
College Park, MD  
Co-instructor (with Prof. Bonnie Dorr) for a computer science graduate seminar on natural language processing with a focus on machine translation literatures.

HARVARD UNIV. DIVISION OF APPLIED SCIENCES Spring 1999  
Cambridge, MA  
Teaching assistant for a graduate seminar on natural language processing taught by Prof. Stuart Shieber. Responsibilities include selecting course reading material, helping students with learning the Prolog programming language, grading assignments as well as handling administrative issues.

**Honors** University of Pittsburgh Nominee for the Microsoft New Faculty Fellowship (2005).  
Dept. of Computer Science Teaching Award, Univ. of Pittsburgh (2004).  
National Science Foundation Graduate Fellowship (1993–1996).  
Outstanding Senior in Computer Science, UCLA (1993).  
Tau Beta Pi Engineering Honors Society.  
UCLA Alumni Scholarship and Achievement Award (1989-1993).

**Professional Activities and Services** Editorial board:  
*Computational Linguistics* (2003-2005).  
Area Chair:  
Joint Conference of Human Language Technologies and the North American Chapter of the ACL (2007).  
The 44th Annual Meeting of the ACL, Student Research Workshop (2006).  
Conference on Empirical Methods in Natural Language Processing (2005).  
Reviewer:  
*Journal of Natural Language Engineering* (2005).  
*Computational Linguistics* (2000, 2002, 2006, 2007).  
*Machine Learning* (2002).  
Program committees:  
International Joint Conference on Artificial Intelligence (2007).  
Conference on Empirical Methods in Natural Language Processing (2001-2004,2007).  
The Annual Meeting of the ACL (2001, 2002, 2003, 2005).  
Selected conference review panels:  
Joint Conference of Human Language Technologies and the North American Chapter of the ACL (2003,2004,2006).  
European Chapter of the ACL (EACL 2003)  
Neural Information Processing Systems (NIPS\*2003), Application Track.  
American Association of Artificial Intelligence (AAAI 2005).

Invited participation:

NSF Workshop: Developing Linguistic Resources with the ANC (2006).

Johns Hopkins Workshop on Language Engineering (2002, 2005).

Language Variation Working Session at the Center for Advanced Study  
of Language, University of Maryland (2004).

Computer Research Association - Women Distributed Mentor Project (2004).

National Science Foundation review panel (2002, 2003, 2004, 2005).

Outreach:

North American Computational Linguistics Olympiad,  
Tutorial session co-chair (2007).

**Personal**

U.S. Citizen, fluent in Mandarin Chinese.