

# REBECCA HWA

Department of Computer Science  
University of Pittsburgh, Pittsburgh, PA 15260, USA  
(412) 624-7953  
[hwa@cs.pitt.edu](mailto:hwa@cs.pitt.edu)  
<http://www.cs.pitt.edu/~hwa>

## EDUCATION

PhD in computer science, 2001, Harvard University (Advisor: Stuart Shieber).

BS in computer science and engineering (summa cum laude), 1993, UCLA.

## PROFESSIONAL EXPERIENCES

Assistant Professor, Department of Computer Science, University of Pittsburgh, PA (2003 – present)

Postdoctoral Fellow, Institute for Advanced Computer Studies, University of Maryland at College Park, MD (2001 – 2003)

Independent Contractor, Applied Language Technologies, Inc. (now a part of Nuance), Cambridge, MA (1995 – 1996).

## HONORS

National Science Foundation CAREER Award (2008).

Department of Computer Science Teaching Award, University of Pittsburgh (2004, 2007).

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).

## RESEARCH, SCHOLARLY, AND CREATIVE ACTIVITIES

### REFEREED JOURNAL ARTICLES

- (J.1) Joshua S. Albrecht and Rebecca Hwa. 2008. Regression for Sentence-Level Machine Translation Evaluation. To be submitted to *Natural Language Engineering*.
- (J.2) Kaihong Liu, Wendy Chapman, Rebecca Hwa, and Rebecca 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*.
- (J.3) Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence* No. 22, Vol. 2.
- (J.4) Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. 2005. Bootstrapping Parsers via Syntactic Projection across Parallel Texts. *Special Issue of the Journal of Natural Language Engineering on Parallel Texts*, Eds. Rada Mihalcea and Michel Simard. No. 11, Vol. 3.
- (J.5) Rebecca Hwa. 2004. Sample selection for statistical parsing. *Computational Linguistics*. No. 30, Vol. 3.
- (J.6) Bonnie J. Dorr, N. Fazil Ayan, Nizar Habash, Nitin Madnani, and Rebecca Hwa, 2003. Rapid Porting of DUSTER to Hindi. In *ACM Transactions on Asian Language Information Processing (TALIP)*, No. 2, Vol. 3.

### REFEREED CONFERENCE PAPERS

- (C.1) Joshua Albrecht and Rebecca Hwa. 2007. A Re-examination of Machine Learning Approaches for Sentence-Level MT Evaluation. In The Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic. [22%]
- (C.2) Joshua Albrecht and Rebecca Hwa. 2007. Regression for Sentence-Level MT Evaluation with Pseudo References. In The Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic. [22%]
- (C.3) Marcus Mylonakis, Khalil Sima'an, and Rebecca Hwa. 2007. Unsupervised Estimation for Noisy-Channel Models. In *The Proceedings of International Conference of Machine Learning (ICML)*, Corvallis, Oregon. [29%]
- (C.4) Rebeca Hwa, Carol Nichols, Khalil 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. [44%]

- (C.5) Chenhai Xi and Rebecca Hwa. 2005. A Backoff Model for Bootstrapping Resources for Non-English Languages. In *The Proceedings of the Joint Conference of Human Language Technologies and Empirical Methods for Natural Language Processing (HLT/EMNLP-05)*, Vancouver, Canada. [32%]
- (C.6) Theresa Wilson, Janyce Wiebe, and Rebecca Hwa. 2004. Just how mad are you? Finding strong and weak opinion clauses. In *The Proceedings of The Nineteenth National Conference on Artificial Intelligence (AAAI-2004)*, San Jose, California. [26.5%]
- (C.7) Mark Steedman, Rebecca Hwa, Stephen Clark, Miles Osborne, Anoop Sarkar, Julia Hockenmaier, Paul Ruhlen, Steven Baker, and Jeremiah Crim. 2003. Example Selection for Bootstrapping Statistical Parsers. In *the Proceedings of the joint Conference of Human Language Technologies and the North American Chapter of the Association for Computational Linguistics (HLT/NAACL-2003)*, Edmonton, Canada. [23%]
- (C.8) Mark Steedman, Miles Osborne, Anoop Sarkar, Stephen Clark, Rebecca Hwa, Julia Hockenmaier, Paul Ruhlen, Steven. Baker, and Jeremiah Crim. 2003. Bootstrapping Statistical Parsers from Small Datasets. In *The Proceedings of the Tenth Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary. [26.5%]
- (C.9) Rebecca Hwa, Philip Resnik, Amy Weinberg, and Okan Kolak. 2002. Evaluating Translational Correspondence using Annotation Projection. In *The Proceedings of the 40<sup>th</sup> Anniversary Meeting of the Association for Computational Linguistics*, Philadelphia, Pennsylvania.[26%]
- (C.10) Bonnie Dorr, Lisa Pearl, Rebecca Hwa, and Nizar 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. [N/A]
- (C.11) Rebecca Hwa. 2001. On Minimizing Training Data for Parser Acquisition. In *The Proceedings of the Conference on Computational Natural Language Learning (CoNLL-2001)*. [33%]
- (C.12) Rebecca 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*, Hong Kong, China. [32%]
- (C.13) Rebecca Hwa. 1999. Supervised Grammar Induction using Training Data with Limited Constituent Information. In *The Proceedings of the 37<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, College Park, Maryland. [25%]

(C.14) Rebecca Hwa. 1998. An Empirical Evaluation of Probabilistic Lexicalized Tree Insertion Grammars. In *the Proceedings of the 36<sup>th</sup> Annual Meeting of the Association for Computational Linguistics and 17<sup>th</sup> International Conference on Computational Linguistics (COLING-ACL 1998)*, volume 1, Montreal, Canada. [25%]

---

## REFEREED WORKSHOP PAPERS

- (W.1) Joshua S. Albrecht, Rebecca Hwa, and G. Elisabeta Marai. 2008. The Chinese Room : Understanding and Correcting Machine Translation. To appear in *The Proceedings of IEEE Visualization 2008 Poster Compendium*.
- (W.2) Joshua Albrecht and Rebecca Hwa, 2008. The Role of Pseudo References in MT Evaluation. In *The Proceedings of the Third Workshop on Statistical Machine Translation at the 46<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Columbus, OH.
- (W.3) Behrang Mohit and Rebecca Hwa. 2007. Localization of Difficult-to-Translate Phrases. In *the Proceedings of the 2nd Workshop on Statistical Machine Translations at the 45<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, Prague, Czech Republic. [32% for oral presentation]
- (W.4) Behrang Mohit and Rebecca Hwa. 2005. Syntax-based Semi-Supervised Named Entity Tagging. In *The Proceedings of the 43<sup>rd</sup> Meeting of the Association for Computational Linguistics (Poster/Demo Session)*, Ann Arbor, Michigan. [55%]
- (W.5) Carol Nichols and Rebecca Hwa. 2005. Word Alignment and Cross-Lingual Resource Acquisition. In *The Proceedings of the 43<sup>rd</sup> Meeting of the Association for Computational Linguistics (Poster/Demo Session)*, Ann Arbor, Michigan. [55%]
- (W.6) Beatriz Maeireizo, Diane Litman, and Rebecca Hwa. 2004. Co-training for Predicting Emotions with Spoken Dialogue Data. In *The Proceedings of the 42<sup>nd</sup> Annual Meeting of the Association for Computational Linguistics Poster/Demo Session*. [45%]
- (W.7) Rebecca Hwa, Miles Osborne, Anoop Sarkar, and Mark 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 at the International Conference of Machine Learning (ICML)*, Washington D.C..
- (W.8) Rebecca Hwa, Philip Resnik, and Amy 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.

- (W.9) Adam Lopez, Michael Nossal, Rebecca Hwa, and Philip 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.

---

## DISSERTATION

Rebecca Hwa. *Learning Probabilistic Lexicalized Grammars for Natural Language Processing*. Harvard University, 2001.

---

## GRANTS AND CONTRACTS

- (G.1) CAREER: Robust Parsing for New Domains and Languages  
Rebecca Hwa  
National Science Foundation
- (G.2) RAVINE: Recombination, Aggregation, and Visualization of Information in News-worthy Expressions  
Noah Smith (PI, CMU), Alan Black (CMU), Frederick Crabbe (USNA), Rebecca Hwa  
Department of Interior
- (G.3) NLP Foundational Studies and Ontologies for Syndromic Surveillance from ED Reports  
Wendy Chapman (PI), Rebecca Hwa, Janyce Wiebe  
National Institute of Health
- (G.4) Collaborative: Discriminative Knowledge-Rich Language Modeling for Machine  
Rebecca Hwa (Co-PI), Alon Lavie (Co-PI, CMU)  
National Science Foundation
- (G.5) SGER: Learning Syntax-based Evaluation Metrics for Machine Translation  
Rebecca Hwa  
National Science Foundation

---

## PATENT

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

## TEACHING

### COURSES TAUGHT

Term	Course
Fall 2003	CS 1590 (UG) Social Implications of Computing
Spring 2004	CS 3730 (G) Advanced Topics in Natural Language Processing
Fall 2004	CS 1571 (UG) Introduction to Artificial Intelligence
Spring 2004	CS 2750 (G) Machine Learning
Fall 2005	CS 1566 (UG) Introduction to Computer Graphics
Fall 2005	CS 1571 (UG) Introduction to Artificial Intelligence
Spring 2006	CS 2750 (G) Machine Learning
Fall 2006	CS 1566 (UG) Introduction to Computer Graphics
Fall 2006	CS 1590 (UG) Social Implications of Computing
Spring 2007	CS 3730 (G) Advanced Topics in Natural Language Processing
Fall 2007	CS 2731 (G) Natural Language Processing
Spring 2008	CS 1671 (UG) Introduction to Human Language Technologies
Spring 2008	CS 2750 (G) Machine Learning

## SERVICE

### PROFESSIONAL ACTIVITIES

#### SELECTED CONFERENCE REVIEW

The 11<sup>th</sup> International Conference on Theoretical and Methodological Issues in Machine Translation (*TMI-07*).

Joint Conference of Human Language Technologies and the North American Chapter of the Association for Computational Linguistics (2003, 2004, 2006).

The Tenth Conference of the European Chapter of the Association for Computational Linguistics (EACL 2003).

The 16<sup>th</sup> Conference on Advances in Neural Information Processing Systems (NIPS\*2003), Application Track.

The Twentieth National Conference on Artificial Intelligence (AAAI 2005).

## **PROGRAM COMMITTEES**

International Joint Conference on Artificial Intelligence (IJCAI), (2007).

Conference on Empirical Methods in Natural Language Processing (2001-2004,2007).

The Annual Meeting of the Association for Computational Linguistics (2001, 2002, 2003, 2005).

## **JOURNAL REVIEW**

Journal of Artificial Intelligence (2008).

Journal of Information Technology & Politics (2007).

Journal of Natural Language Engineering (2005).

Computational Linguistics (2000, 2002, 2006, 2007).

Machine Learning (2002).

---

## **OUTREACH**

Technology Leadership Initiative – Lecture presentations (2006, 2008)

North American Computational Linguistics Olympiad – Tutorial Co-chair (2007).

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

## **PROFESSIONAL ACTIVITIES**

---

### **INVITED TALKS**

- “Learning Evaluation Metrics for Sentence-Level Machine Translation”
  - Cornell University, NY, April 2008.
  - Carnegie Mellon University, PA, September 2007.
- “Building Resources for Parsing Arabic Dialects”
  - Carnegie Mellon University, PA, October 2005.
- “Semi-supervised Learning for Information Management”
  - University of Maryland, College Park. MD, 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.

---

## EXECUTIVE BOARD

North American Chapter of the Association for Computational Linguistics – Board Member (2008 – 2010).

---

## EDITORIAL BOARD

Computational Linguistics (2003-2005).

---

## AREA CHAIR

The Conference on Empirical Methods in Natural Language Processing, Parsing Track (2008).

Joint Conference of Human Language Technologies and the North American Chapter of the Association for Computational Linguistics, Machine Translation Track (2007).

The 44th Annual Meeting of the ACL, Student Research Workshop, Faculty Advisor (2006).

The Joint Conference of Human Language Technologies and Empirical Methods in Natural Language Processing, Parsing Track (2005).

---

## INVITED PARTICIPATION

National Science Foundation Panel: Scientific Community Onsite Assessment Workshop (2008).

National Science Foundation Workshop Panel: Developing Linguistic Resources with the American National Corpus (2006).

National Science Foundation Review Panel (2002, 2003, 2004, 2005).

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

University of Maryland Center for Advanced Study of Language Panel: Language Variation Working Session (2004).