

# 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

Associate Professor, Department of Computer Science, University of Pittsburgh, PA (2009 – present)

Assistant Professor, Department of Computer Science, University of Pittsburgh, PA (2003 – 2009)

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 Undergraduate in Computer Science, UCLA (1993).

### REFEREED JOURNAL ARTICLES

- (J.1) Joshua S. Albrecht, Rebecca Hwa, and G. Elisabeta. Marai. The Chinese Room: Visualization and Interaction to Understand and Correct Ambiguous Machine Translation. *Computer Graphics Forum*, Vol. 28, No. 3.
- (J.2) Joshua S. Albrecht and Rebecca Hwa. 2008. Regression for Sentence-Level Machine Translation Evaluation. Submitted to *Machine Translation*.
- (J.3) 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.4) Theresa Wilson , Janyce Wiebe, and Rebecca Hwa. 2006. Recognizing strong and weak opinion clauses. *Computational Intelligence* No. 22, Vol. 2.
- (J.5) 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.6) Rebecca Hwa. 2004. Sample selection for statistical parsing. *Computational Linguistics*. No. 30, Vol. 3.
- (J.7) 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) Huichao Xue and Rebecca Hwa. 2014. Redundancy Detection in ESL Writings. In *The Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, Gothenburg, Sweden.
- (C.2) Huichao Xue and Rebecca Hwa. 2012. Modeling ESL Word Choice Similarities By Representing Word Intensions and Extensions. In *The Proceedings of the 24th International Conference on Computational Linguistics (COLING-2012)*, Mumbai, India.
- (C.3) Matthias Grabmair, Kevin D. Ashley, Rebecca Hwa, Patricia M. Sweeney. 2011. Toward Extracting Information from Public Health Statutes using Text Classification Machine Learning. In *the Proceedings of the 24<sup>th</sup> International*

*Conference on Legal Knowledge and Information Systems (Jurix-2011)*,  
Vienna, Austria.

- (C.4) Huichao Xue and Rebecca Hwa. 2010. Syntax-Driven Machine Translation as a Model of ESL Revision. In *The Proceedings of the 23th International Conference on Computational Linguistics (COLING-2010)*, Beijing, China.
- (C.5) Joshua S. Albrecht, Rebecca Hwa, and G. Elisabeta Marai. 2009. Correcting Automatic Translations through Collaborations between MT and Monolingual Target-Language Users. In *The Proceedings of the 12th Conference of the European Chapter of the ACL (EACL-2009)*, Athens, Greece.
- (C.6) Behrang Mohit, Frank Liberato, and Rebecca Hwa. 2009. Language Model Adaptation for Difficult-to-Translate Phrases. In *The Proceedings of the 13th Annual Conference of the European Association for Machine Translation (EAMT-09)*, Barcelona, Spain.
- (C.7) 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.
- (C.8) 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.
- (C.9) 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.
- (C.10) 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.
- (C.11) 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.
- (C.12) 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.
- (C.13) 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.

- (C.14) 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.
- (C.15) 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.
- (C.16) 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.
- (C.17) Rebecca Hwa. 2001. On Minimizing Training Data for Parser Acquisition. In *The Proceedings of the Conference on Computational Natural Language Learning (CoNLL-2001)*.
- (C.18) 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.
- (C.19) 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.
- (C.20) 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.

---

#### REFEREED SHORT PAPERS, WORKSHOP PAPERS, AND DEMO PAPERS

- (W.1) Huichao Xue and Rebecca Hwa. 2014. Improved Correction Detection in Revised ESL Sentences. In *The Proceedings of The 52nd Annual Meeting of the Association for Computational Linguistics (ACL short)*, 2014, Baltimore, Maryland.
- (W.2) Sisi Lu, Ye Ye, Rich Tsui, Xiaochu Liu, Rebecca Hwa. *Feature Selection for 30-Day Heart Failure Read-mission Prediction Using Clinical Drug Data*. NIPS Workshop on Machine Learning for Clinical Data Analysis and Healthcare, 2013.

- (W.3) Homa Hashemi and Rebecca Hwa. 2014. A Comparison of MT Errors and ESL Errors. In the Proceedings of the 9th edition of the Language Resources and Evaluation Conference (LREC-2014), Reykjavik, Iceland.
- (W.4) Phuong Pham, Janyce Wiebe, Rebecca Hwa, Wendy Chapman. 2013. Automated Annotation of Colonoscopy Reports. *AMIA Summit on Clinical Research Informatics (CRI 2013)*, San Francisco, California.
- (W.5) Alexander Conrad, Harry Hochheiser, Wendy Chapman, Rebecca Hwa, and Janyce Wiebe. 2013. Tightly-Coupled Natural Language Processing and Visualization for Electronic Medical Record Chart Review. *AMIA Summit on Clinical Research Informatics (CRI 2013)* San Francisco, California.
- (W.6) Alexander Conrad, Janyce Wiebe, and Rebecca Hwa. 2012. Recognizing Arguing Subjectivity and Argument Tags. In *the Proceedings of the Workshop on Extra-Propositional Aspects of Meaning in Computational Linguistics (ExProM '12)*, Jeju Island, Korea.
- (W.7) Frank Liberato, Behrang Mohit, and Rebecca Hwa. 2010. Improving Phrase-Based Translation with Prototypes of Short Phrases. In *The Proceedings of the 11<sup>th</sup> Annual Conference of the North American Chapter of the Association for Computational Linguistic*, Los Angeles.
- (W.8) Joshua S. Albrecht, Rebecca Hwa, and G. Elisabeta Marai. 2008. The Chinese Room : Understanding and Correcting Machine Translation. In *The Proceedings of IEEE Visualization 2008 Poster Compendium*.
- (W.9) 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.10) Noah Smith, Michael Heilman, and Rebecca Hwa, 2008. Question Generation as a Competitive Undergraduate Course Project. In *The Proceedings of the NSF Workshop on the Question Generation Shared Task and Evaluation Challenge*, Arlington VA.
- (W.11) 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.
- (W.12) 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.

- (W.13) 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.
- (W.14) 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*.
- (W.15) 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.16) 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.17) 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) R01: Interactive Search and Review of Clinical Records with Multi-layered Semantic Annotation  
Wendy Chapman (PI), Rebecca Hwa, Janyce Wiebe  
National Institute of Health  
Amount awarded for Pitt's subcontract: \$940,000 for four years, starting October 2011.
- (G.2) CAREER: Robust Parsing for New Domains and Languages  
Rebecca Hwa  
National Science Foundation  
Amount awarded: \$499,984 for five years, starting July 2008.

- (G.3) 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  
Amount awarded for Pitt's subcontract: \$48,689 for two years, starting April 2008.
- (G.4) NLP Foundational Studies and Ontologies for Syndromic Surveillance from ED Reports  
Wendy Chapman (PI), Rebecca Hwa, Janyce Wiebe  
National Institute of Health  
Amount awarded: Total of \$1,257,457 for three years, starting August 2007.
- (G.5) Collaborative: Discriminative Knowledge-Rich Language Modeling for Machine  
Rebecca Hwa (Co-PI), Alon Lavie (Co-PI, CMU)  
National Science Foundation  
Amount awarded for the Pitt side: \$87,389 for three years, starting July 2007; and \$6,000 for REU Supplemental in 2008.
- (G.6) SGER: Learning Syntax-based Evaluation Metrics for Machine Translation  
Rebecca Hwa  
National Science Foundation  
Amount awarded: \$194,344 for two years, starting July 2006; and \$11,113 for two REU Supplementals, in 2007.

---

## 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.

## SERVICE

---

### ON-CAMPUS COMMITTEES

School of Arts and Sciences Undergraduate Research Awards Selection Committee (2006 – 2008)

School of Arts and Sciences Honors College Advisory Board (2009-2011)

Department of Computer Science Faculty Search Committee (2010)

Department of Computer Science Graduate Admissions & Financial Aid Committee (2008-2010)

Department of Computer Science Undergraduate Advising Committee (2003 – 2008)

Department of Computer Science Graduate Program & Examinations Committee (2003 – 2005)

---

## OUTREACH

Linguistic Institute: Statistical Syntactic Parsing – Instructor (2011)

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

- “Relationships between Machine Translation and Grammar Error Correction”
  - IBM Watson, NY, March 2014.
- “The Role of Machine Translation in Modeling English as a Second Language (ESL) Writings”
  - Universite de Montreal, Canada, February 2012.
  - University of Maryland, MD, March 2012.
- “Applications of information visualization for natural language processing,”
  - Johns Hopkins University, MD, April 2011.
  - University of Illinois at Urbana Champagne, IL, October 2010.
  - Carnegie Mellon University, PA, May 2010.
- “Learning Evaluation Metrics for Sentence-Level Machine Translation”
  - University of Texas, Austin, October 2008.
  - 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.



---

## 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, 2010, 2013).

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

---

## OFFICES

North American Chapter of the Association for Computational Linguistics – Past Chair of the Executive Board (2012 – 2014).

North American Chapter of the Association for Computational Linguistics – Chair of the Executive Board (2010 – 2012).

North American Chapter of the Association for Computational Linguistics – Executive board member (2008 – 2010).

Computational Linguistics – Editorial board member (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).

---

## SELECTED PROGRAM COMMITTEES

The Annual Meeting of the Association for Computational Linguistics (2001, 2002, 2003, 2005, 2008, 2009, 2010, 2013, 2014).

The Annual Meetings of the North American Chapter of the Association for Computational Linguistics (2003, 2004, 2006, 2009, 2010, 2013).

The Conferences of the European Chapter of the Association for Computational Linguistics (2003, 2009, 2014).

Conference on Empirical Methods in Natural Language Processing (2001-2004, 2007, 2013, 2014).

International Conference on Computational Linguistics (2010)

Conference on Computational Natural Language Learning (2010, 2013, 2014)

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).

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

---

## JOURNAL REVIEW

Transaction of the Association for Computational Linguistics (2013, 2014)

Journal of Selected Topics in Signal Processing (2009)

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).