

## YEJIN CHOI

Cornell University  
Upson Hall  
Ithaca, NY 14853

<http://www.cs.cornell.edu/~ychoi>

ychoi@cs.cornell.edu  
607-280-4546

Citizenship: South Korea

---

### RESEARCH INTERESTS:

- Statistical approaches and computational models for Natural Language Processing and Computational Linguistics
- Interdisciplinary research connecting Natural Language Processing with Psycholinguistics, Sociolinguistics, Psychology, and Cognitive Science

### EDUCATION:

- **Cornell University, Ithaca, NY, USA**  
Ph.D. in Computer Science (Expected in May 2010)  
Advisor: Claire Cardie  
  
M.S in Computer Science (Mar 2009)  
  
Minor in Operations Research (Expected in May 2010)  
Minor advisor: David P. Williamson
- **Seoul National University, Seoul, Korea**  
B.S. with honor in Computer Science and Engineering (Feb 1999)

### EXPERIENCE:

- **Yahoo! Research**, Santa Clara, CA (Jun 2009 – Sep 2009)  
**Research Intern** for Computational Advertising Group  
Advisor: Vanja Josifovski, Marcus F. Fontoura
- **Cornell University**, Ithaca, NY (May 2004 – present)  
**Research Assistant**
  - ❖ Part of the system I built for opinion analysis has been employed by Jodange ( <http://www.jodange.com> ), a start-up company co-founded by my advisor Claire Cardie.
- **Microsoft Corporation**, Redmond, WA (May 2000 – Jul 2003)  
**Software Design Engineer** for Windows Media Platform Group

- **LG Electronics Research Center**, Seoul, Korea (May 1999 – Mar 2000)  
Research Staff for HDTV data broadcast software solution team

#### HONORS:

- Outstanding TA award, Cornell University, 2004

#### TEACHING EXPERIENCE:

- **Guest Lecturer**  
Advanced Language Technologies (Fall 2009)  
Introduction to Natural Language Processing (Spring 2009)  
Operating Systems (Fall 2003)
- **Teaching Assistant**  
Introduction to Natural Language Processing (Spring 2009, Spring 2006, Spring 2005)  
Java Practicum (Spring 2004)  
Operating Systems (Fall 2003)

#### PUBLICATIONS:

##### - CONFERENCE (REFEREED) –

- Using Landing Pages for Sponsored Search Ad Selection  
Yejin Choi, Marcus Fontoura, Evgeniy Gabrilovich, Vanja Josifovski, Mauricio Mediano, and Bo Pang  
International World Wide Web Conference (WWW), 2010.
- Adapting a Polarity Lexicon Using Integer Linear Programming for Domain-Specific Sentiment Classification.  
Yejin Choi, and Claire Cardie.  
Empirical Methods in Natural Language Processing (EMNLP), 2009.
- Learning with Compositional Semantics as Structural Inference for Subsentential Sentiment Analysis.  
Yejin Choi, and Claire Cardie.  
Empirical Methods in Natural Language Processing (EMNLP), 2008.
- Structured Local Training and Biased Potential Functions for Conditional Random Fields with Application to Coreference Resolution  
Yejin Choi, and Claire Cardie.  
Human Language Technologies: the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2007.

- Identifying Expressions of Opinion in Context.  
Eric Breck, Yejin Choi, and Claire Cardie.  
International Joint Conference on Artificial Intelligence (IJCAI), 2007.
- Joint Extraction of Entities and Relations for Opinion Recognition.  
Yejin Choi, Eric Breck, and Claire Cardie.  
Empirical Methods in Natural Language Processing (EMNLP), 2006.
- Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns.  
Yejin Choi, Claire Cardie, Ellen Riloff, and Siddharth Patwardhan.  
Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing (EMNLP), 2005.

**- WORKSHOP & DEMO (REFEREED) –**

- Cornell System Description for the NTCIR-6 Opinion Task.  
Eric Breck, Yejin Choi, Veselin Stoyanov, and Claire Cardie.  
The 6th NTCIR Workshop Meeting , 2007.
- OpinionFinder: A System for Subjectivity Analysis.  
Theresa Wilson, Paul Hoffmann, Swapna Somasundaran, Jason Kessler, Janyce Wiebe, Yejin Choi, Claire Cardie, Ellen Riloff and Siddharth Patwardhan.  
A demo at Human Language Technology Conference/Conference on Empirical Methods in Natural Language Processing (EMNLP), 2005.

**- CONFERENCE (IN SUBMISSION) –**

- Automatically Generating Annotator Rationales to Improve Sentiment Classification.  
Ainur Yessenalina, Yejin Choi, and Claire Cardie

**- UNPUBLISHED MANUSCRIPTS –**

- Hierarchical Sequential Learning for Extracting Opinions and their Attributes.  
Yejin Choi, and Claire Cardie

**- PATENTS –**

- System and method for summarizing fine-grained opinions in text  
C. Cardie, Y. Choi, E. Breck, V. Stoyanov  
[Patent pending] filed by Jodange, LLC, 2007.

- Sparse caching for streaming media.  
A. Grigorovitch, Y. Choi , T. Carvalho.  
[Patent 20050066063] filed by Microsoft Corporation, 2003
- Midstream determination of varying bandwidth availability.  
T. Batterberry, A. Grigorovitch, A. Klemets, J. Stewart, Y. Choi  
[Patent 20040267503] filed by Microsoft Corporation, 2003.
- System and method for automatically recovering from failed network connections in streaming media scenarios.  
Y. Choi , A. Grigorovitch, T. Batterberry  
[Patent 20030236905] filed by Microsoft Corporation, 2002.

#### **PROFESSIONAL ACTIVITIES:**

- Program Committee for
  - TextGraphs Workshop at ACL (2010)
  - Association for Computational Linguistics (ACL) (2009, 2010)
  - International AAAI Conference on Weblogs and Social Media (ICWSM) (2010)
- Reviewer for
  - Association for Computational Linguistics (ACL) (2006, 2007, 2008, 2009)
  - Empirical Methods on Natural Language Processing (EMNLP) (2007, 2008, 2009)
  - Neural Information Processing Systems (NIPS) (2009)
  - North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT) (2010)
  - Language Resources and Evaluation (LRE) Journal (2009)
  - IEEE Transactions on Audio, Speech and Language Processing (2007)
  - Knowledge and Information Systems Journal (2005)

#### **SKILLS:**

- Course work at Cornell: (GPA: 4.03)
  - Computer Science -
    - Learning to predict structured objects, Advanced algorithms, Advanced artificial intelligence, Advanced database systems, Advanced programming languages, Natural language processing, Reasoning about knowledge, Statistical learning theory
  - Operations Research -
    - Stochastic processes, Approximation algorithms, Statistical principles

- Language Proficiency
  - English    Fluent knowledge
  - Korean    Mother tongue
  - Polish     Basic knowledge

#### REFERENCES:

- Claire Cardie  
Professor  
Cornell University

Department of Computer Science  
5161 Upson Hall  
Ithaca, NY 14853  
Phone: 607-255-9206  
Fax: 607-255-4428  
Email: [cardie@cs.cornell.edu](mailto:cardie@cs.cornell.edu)

- Lillian Lee  
Professor  
Cornell University

Department of Computer Science  
4152 Upson Hall  
Ithaca, NY 14853  
Phone: 607-255-8119  
Fax: 607-255-4428  
Email: [llee@cs.cornell.edu](mailto:llee@cs.cornell.edu)

- Ellen Riloff  
Associate Professor  
University of Utah

School of Computing  
50 S. Central Campus Drive, Rm  
3190 MEB  
Salt Lake City, UT 84112-9205  
Phone: 801-581-7544  
Fax: 801-581-5843  
Email: [riloff@cs.utah.edu](mailto:riloff@cs.utah.edu)

- Vanja Josifovski  
Principal Research Scientist  
Yahoo! Research

4301 Great America Parkway  
Santa Clara, CA 95054  
Email: [vanjaj@yahoo-inc.com](mailto:vanjaj@yahoo-inc.com)