

Kristina N. Toutanova

Microsoft Research,
One Microsoft Way,
Redmond, WA 98052
(425) 705-3259
kristout@microsoft.com

EDUCATION

- 1999-2005* **Stanford University, Computer Science Department**
Ph.D. received September 2005
Advisor: Christopher Manning
Dissertation Title: "Effective Probabilistic Models for Syntactic and Semantic Disambiguation"
- 1998-1999* **Sofia University, Bulgaria, Computer Science Department**
Master of Science, Concentration in Artificial Intelligence
- 1994-1998* **Sofia University, Bulgaria, Computer Science Department**
Bachelor of Science, Computer Science
GPA 6.0/6.0
- 1997* **University College London, Computer Science Department**
One semester study

RESEARCH INTERESTS

Natural Language Processing (Machine Translation, Morphology, Parsing, Question Answering, Information Extraction), Machine Learning, Information Retrieval, AI

OTHER RESEARCH EXPERIENCE

- 2005-present* **Microsoft Research, Redmond**
Researcher
Performing research in the areas of machine translation and machine learning for natural language processing.
- 2001 summer* **Microsoft Research, Redmond**
Research Intern
Implemented a statistical model for automatic spelling correction, which combines letter-based and phone-based knowledge.

- 2000 summer* **Xerox Palo Alto Research Center**
Research Intern
Built hierarchical probabilistic models for document topic analysis.
- 1998 December* **Hamburg University, Computer Science**
Visiting Research Student
- 1996-1999* **Bulgarian Academy of Sciences, Linguistic Modelling Department**
Research Programmer
Performed research in natural language generation and knowledge representation.

TEACHING EXPERIENCE

- 2003 Spring* **Teaching Assistant for CS224n, Natural Language Processing, Stanford University**
- 2002 Autumn* **Teaching Assistant for CS221, Artificial Intelligence, Stanford University**
- 2001 Winter* **Teaching Assistant for CS121, Artificial Intelligence, Stanford University**
- 1997* **Demonstrator for Mathematics for Economics Students, University College London**

AWARDS AND HONORS

- 2007* Technical Innovation Award at the Microsoft Women's Leadership and Development Conference
- 2003* Best student paper runner up award at ECML-2003 "Optimizing Local Probability Models for Statistical Parsing"
- 1999-2000* School of Engineering Fellowship, Stanford University
- 1994-1999* Ministry of Education scholarship for excellence in academics, Sofia University

PUBLICATIONS

Journal Articles

Kristina Toutanova, Aria Haghighi, and Christopher D. Manning. 2008. in press. A Global Joint Model for Semantic Role Labeling. *Computational Linguistics* 34(2) (Special Issue on Semantic Role Labeling).

Kristina Toutanova, Christopher D. Manning, Dan Flickinger, Stephan Oepen. 2005. “Stochastic HPSG Parse Disambiguation using the Redwoods Corpus”. *Journal of Research on Language and Computation*, 3(1):83-105, 2005.

Stephan Oepen, Daniel Flickinger, Kristina Toutanova, and Christopher D. Manning. LinGO Redwoods. A rich and dynamic treebank for HPSG. *Journal of Research on Language and Computation*, 2(4):575-596, 2004

Refereed Conference Papers

Xiaodong He and Kristina Toutanova. 2009. “Joint Optimization for Machine Translation System Combination”. *Proceedings of the Conference of Empirical Methods in Natural Language Processing, 2009*.

Kristina Toutanova and Colin Cherry. 2009. “A Global Model for Joint Lemmatization and Part-of-speech Prediction”. *Proceedings of the 47th Meeting of the Association for Computational Linguistics, 2009*.

Hoifung Poon, Colin Cherry, and Kristina Toutanova. 2009. “Unsupervised Morphological Segmentation with Log-Linear Models”. *Proceedings of NAACL-HLT, 2009*. **Best paper award.**

Jia Xu, Jianfeng Gao, Kristina Toutanova, and Hermann Ney. 2008. “Bayesian Semi-Supervised Chinese Word Segmentation for Statistical Machine Translation”. *Proceedings of the 22nd International Conference on Computational Linguistics, 2008*.

Kristina Toutanova, Hisami Suzuki, and Achim Ruopp. 2008. “Applying Morphology Generation Models to Machine Translation”. *Proceedings of the 46th Meeting of the Association for Computational Linguistics, 2008*.

Kristina Toutanova and Mark Johnson. 2007. A Bayesian LDA-based Model for Semi-supervised Part-of-Speech Tagging. In *Proceedings of NIPS 21, 2007*.

Pi-Chuan Chang and Kristina Toutanova. 2007. “A Discriminative Syntactic Word Order Model for Machine Translation”. *Proceedings of the 45th Meeting of the Association for Computational Linguistics, 2007*.

Einat Minkov, Kristina Toutanova, and Hisami Suzuki. 2007. “Generating Complex Morphology for Machine Translation”. *Proceedings of the 45th Meeting of the Association for Computational Linguistics, 2007*.

Kristina Toutanova and Hisami Suzuki. 2007. “Generating Case Markers in Machine Translation”. *Proceedings of the Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2007*.

Kristina Toutanova. 2006. “Competitive Generative Models with Structure Learning for NLP Classification Tasks”. *Proceedings of the Conference of Empirical Methods in Natural Language Processing, 2006*.

- Hisami Suzuki and Kristina Toutanova. 2006. "Learning to Predict Case Markers in Japanese". *Proceedings of the Joint Conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics (ACL-COLING), 2006*.
- Aria Haghighi, Kristina Toutanova, and Christopher D. Manning. 2005. "A Joint Model for Semantic Role Labeling". *Proceedings of the Ninth Conference on Natural Language Learning Shared Task 2005*.
- Kristina Toutanova, Aria Haghighi, and Christopher D. Manning. 2005. "Joint learning improves semantic role labeling". *Proceedings of the 43rd Meeting of the Association for Computational Linguistics, 2005*.
- Kristina Toutanova, Penka Markova, and Christopher D. Manning. 2004. "The Leaf Projection Path View of Parse Trees: Exploring String Kernels for HPSG Parse Selection". *Proceedings of the Conference of Empirical Methods in Natural Language Processing, 2004*.
- Kristina Toutanova, Christopher D. Manning, and Andrew Y. Ng. 2004. "Learning Random Walk Models for Inducing Word Dependency Distributions". *Proceedings of the 21st International Conference on Machine Learning, 2004*.
- Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network". *Proceedings of the North American Chapter of the Association for Computational Linguistics, 2003*.
- Kristina Toutanova, Mark Mitchell and Christopher D. Manning. 2003. "Optimizing Local Probability Models for Statistical Parsing". *Proceedings of the European Conference on Machine Learning, 2003*. [Best student paper runner-up award]
- Kristina Toutanova and Robert C. Moore. 2002. "Pronunciation Modeling for Improved Spelling Correction". *Proceedings of the 40th Meeting of the Association for Computational Lstics, 2002*.
- Kristina Toutanova, H. Tolga Ilhan, and Christopher D. Manning. 2002. "Extensions to HMM-based Statistical Word Alignment Models". *Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2002*.
- Kristina Toutanova and Christopher D. Manning. 2002. "Feature Selection for a Rich HPSG Grammar Using Decision Trees". *Proceedings of the Sixth Conference on Natural Language Learning, 2002*.
- Stephan Oepen, Kristina Toutanova, Stuart Shieber, Christopher D. Manning, Dan Flickinger, and Thorsten Brants. 2002. "The LinGO Redwoods Treebank: Motivation and Preliminary Applications". *Proceedings of the 19th International Conference on Computational Linguistics, 2002*.
- Kristina Toutanova, Francine Chen, Kris Popat, and Thomas Hofmann. 2001. "Text Classification in a Hierarchical Mixture Model for Small Training Sets". *Proceedings of the Tenth International Conference on Information and Knowledge Management, 2001*.
- Pavlin Dobrev and Kristina Toutanova. 2001. "CGWorld – Architecture and Features". *Proceedings of the International Conference on Conceptual Structures, 2001*.
- Kristina Toutanova and Christopher D. Manning. "Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger". *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, 2000*.
- Svetlana Damianova and Kristina Toutanova. 1999. "Using Conceptual Graphs to Solve a Resource Allocation Task". *Proceedings of the International Conference on Conceptual Structures, 1999*.
- Galia Angelova, Svetlana Damianova, Kristina Toutanova and Kalina Bontcheva. 1997. "Menu-Based Interfaces to Conceptual Graphs – the CGLex Approach". *Proceedings of the Int Conf on Conceptual Structures, 1997*.

Workshop Papers

Kristina Toutanova, Chris Brockett, Michael Gamon, Jagadeesh Jagarlamudi, Hisami Suzuki and Lucy Vanderwende. 2007. "The Pythy Summarization System: Microsoft Research at DUC 2007". *Proceedings of the Document Understanding Conference, 2007*.

Arul Menezes, Kristina Toutanova, and Chris Quirk. 2006. "Microsoft Research Treelet Translation System: NAACL 2006 Europarl Evaluation". *Proceedings of the Workshop on Statistical Machine Translation at NAACL 2006*.

Dan Klein, Kristina Toutanova, Husrev Tolga Ilhan, S. D. Kamvar, and Christopher D. Manning. 2002. "Combining Heterogeneous Classifiers for Word-Sense Disambiguation". *Proceedings of Workshop on Word Sense Disambiguation: Recent Successes and Future Directions at ACL 40, 2002*.

TUTORIALS

Wen-Tau Yih and Kristina Toutanova. 2007. Automatic Semantic Role Labeling. *AAAI-2007 Tutorial*.

Wen-Tau Yih and Kristina Toutanova. 2006. Automatic Semantic Role Labeling. *HLT-NAACL 2006 Tutorial*.

PRODUCT CONTRIBUTIONS AT MICROSOFT

A model for improving the grammaticality of translation for machine translation from a morphology-poor to a morphology-rich language. 2008.

An order model for machine translation. 2008.

Seven patents filed 2006-2008.

INVITED TALKS

"*A Bayesian LDA-based Model for Semi-supervised Part-of-Speech Tagging*" University of Washington, Feb 2008.

"*Generating Morphologically Rich Languages in Machine Translation*" (with Hisami Suzuki)
University of Washington/Microsoft Symposium in Computational Linguistics, Feb. 2007

"*Effective Probabilistic Models for Syntactic and Semantic Disambiguation*"
University of Washington, Computer Science Department, June 2005
Ohio State University, Computer Science Department, March 2005
Carnegie Mellon University, LTI, April 2005
Northwestern University, Computer Science Department, April 2005
University of Texas, Dallas, Computer Science Department, April 2005
University of Wisconsin Madison, Computer Science Department, April 2005
Microsoft Research, March 2005
Google Inc, March 2005
IBM Research, April 2005

"*Automatic Semantic Role Labeling*"
Yahoo Inc, December 2004

“Generative and Discriminative Models in NLP: A Survey”
Stanford Seminar on Computational Learning and Adaptation, November 2002.

PROFESSIONAL SERVICE

Program Chair: CoNLL 2008 PC co-chair

Area Chair: NAACL 2009, area chair for machine learning

Conference Reviewing: NIPS 2009, ACL 2009, EMNLP 2009, IJCAI 2009, ACL 2008, ICML 2008, COLING 2008, AAI 2008, ACL 2007, EMNLP 2007, HLT-NAACL 2007, RANLP 2007, Text Graphs Workshop 2007, NIPS 2007, ACL-COLING 2006, EMNLP 2006, CONLL 2006, RANLP 2005, RANLP 2003

Program Committee: Workshop on Feature Engineering at ACL 2005, RANLP, CG Tools Workshop at ICCS 2001

Software: Part-of-speech tagger available at: <http://nlp.stanford.edu/software/tagger.shtml>

LANGUAGES

Bulgarian: native, *English:* fluent, *Russian:* good