Kristina N. Toutanova

Google Research
601 N 34th St, Seattle, WA 98103
kristout@google.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

1997 **University College London, Computer Science Department**
One semester study

RESEARCH INTERESTS

Natural Language Processing (Information Extraction, Knowledge Base Completion, Machine Translation, Question Answering, Summarization, Grammatical Error Correction, Morphology, Parsing), **Machine Learning (**Structure Prediction, Graphical Models, Neural Models, Topic Models**), Information Retrieval, AI**

RESEARCH EXPERIENCE

2017-current **Google Research Seattle**
Researcher Scientist
Performing research in natural language understanding.

2005-2016 **Microsoft Research, Redmond**
Researcher
Performing research in machine learning for natural language processing, with a recent focus on relation extraction, knowledge base completion, grammatical error correction, machine translation, and summarization.

2010-present **University of Washington, Department of Linguistics**
Affiliate Assistant Professor

2001 summer **Microsoft Research, Redmond**
Research Intern
Implemented a statistical model for automatic spelling correction

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

2011 Spring	Lecturer for LING 575, Statistical Machine Translation, University of Washington
2010 Spring	Lecturer for LING 575, Statistical Machine Translation, University of Washington
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

2013	Honorable mention award at EMNLP 2013 "Regularized Minimum Error Rate Training"
2011	Best paper award at CONLL 2011 "Learning Discriminative Projections for Text Similarity Measures"
2009	Best paper award at NAACL 2009 "Unsupervised Morphological Segmentation with Log-Linear Models"
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

Sauleh Eetemadi, William Lewis, Kristina Toutanova, and Hayder Radha. 2015. Survey of Data-selection Methods in Statistical Machine Translation. *Machine Translation* 29(3) :189-223, 2015.

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

Kristina Toutanova, Chris Brockett, Ke M. Tran, and Saleema Amershi. 2016. " A Dataset and Evaluation Metrics for Abstractive Sentence and Paragraph Compression". *Proceedings of EMNLP*, 2016.

Kristina Toutanova, Xi Victoria Lin, Scott Wen-tau Yih, Hoifung Poon, and Chris Quirk. 2016. "Compositional Learning of Embeddings for Relation Paths in Knowledge Bases and Text". *Proceedings of ACL*, 2016.

Yuval Marton and Kristina Toutanova. E-TIPSY: Search Query Corpus Annotated with Entities, Term Importance, POS Tags, and Syntactic Parses". *Proceedings of LREC*, 2016.

Rebecca Mason, Benjamin Gaska, Pallavi Choudhury, Kristina Toutanova, Ted Hart, Benjamin Van Durme, Bill Dolan and Margaret Mitchell. 2016. "Microsummarization of Online Reviews: An Experimental Study". *Proceedings of AAAI*, 2016.

Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon. 2015. "Representing Text for Joint Embedding of Text and Knowledge Bases". *Proceedings of EMNLP*, 2015.

Kristina Toutanova, Waleed Ammar, Pallavi Choudhury, and Hoifung Poon. 2015. "Model Selection for Type-

supervised Learning with Application to POS Tagging". *Proceedings of CoNLL* 2015.

Ankur P. Parikh, Hoifung Poon, and Kristina Toutanova. 2015. "Grounded Semantic Parsing for Complex Knowledge Extraction". *Proceedings of NAACL,* 2015.

Hoifung Poon, Kristina Toutanova, and Chris Quirk. 2015. "Distant Supervision for Cancer Pathway Extraction from Text". *Proceedings of Pacific Symposium on Biocomputing,* 2015.

Sauleh Eetemadi and Kristina Toutanova. 2014. "Asymmetric Features of human Generated Translation". *Proceedings of EMNLP* 2014.

Avneesh Saluja, Hany Hassan, Kristina Toutanova, and Chris Quirk. 2014. "Graph-based Semi-Supervised Learning of Translation Models from Monolingual Data". *Proceedings of ACL*, 2014.

Michel Galley, Chris Quirk, Colin Cherry, and Kristina Toutanova. 2013 "Regularized Minimum Error Rare Training". *Proceedings of EMNLP* 2013. **Honorable mention award.**

Kristina Toutanova and Byung-Gyu Ahn. 2013. "Learning Non-linear Features for Machine Translation Using Gradient Boosting Machines". *Proceedings of ACL*, 2013.

Hui Zhang, Kristina Toutanova, Chris Quirk, and Jianfeng Gao. 2013. "Beyond Left-to-right: Multiple Decomposition structures for SMT". *Proceedings of NAACL,* 2013.

Chris Quirk, Pallavi Choudhury, Jianfeng Gao, Hisami Suzuki, Kristina Toutanova, Michael Gamon, Wen-tau Yih, Lucy Vanderwende, and Colin Cherry. 2012. "MSR SPLAT, a Language Analysis Toolkit". *Proceedings of NAACL-HLT 2012 Demonstrations*.

Sungchul Kim, Kristina Toutanova, and Hwanjo Yu. 2012. "Multilingual Named Entity Recognition using Parallel Data and Metadata from Wikipedia". *Proceedings of ACL*, 2012.

Jianfeng Gao, Kristina Toutanova, and Wen-tau Yih. 2011. "Clickthrough-Based Latent Semantic Models for Web Search". *Proceedings of SIGIR*, 2011.

Wen-tau Yih, Kristina Toutanova, John Platt, and Chris Meek. 2011. "Learning Discriminative Projections for Text Similarity Measures". *Proceedings of CoNLL*, 2011. **Best paper award.**

Kristina Toutanova and Michel Galley. 2011. "Why Initialization Matters for IBM Model 1: Multiple Optima and Non-Strict Convexity". *Proceedings of ACL*, 2011.

Jason Naradowsky and Kristina Toutanova. 2011. "Unsupervised Bilingual Morpheme Segmentation and Alignment with Context-rich Hidden Semi-Markov Models". *Proceedings of ACL*, 2011.

John C. Platt, Kristina Toutanova, and Scott Wen-Tau Yih. 2010. "Translingual Document Representations from Discriminative Projections". *Proceedings of EMNLP*, 2010.

Minwoo Jeong, Kristina Toutanova, Hisami Suzuki, and Chris Quirk. 2010. "A Discriminative Lexicon Model for Complex Morphology." *Proceedings of AMTA, 2010*.

Jason Smith, Chris Quirk, and Kristina Toutanova. 2010. "Extracting Parallel Sentences from Comparable Corpora using Document Level Alignment." *Proceedings of NAACL, 2010*.

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 Linguistics, 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 and Danqi Chen. 2015. "Observed Versus Latent Features for Knowledge Base and Text Inference". *Proceedings of 3rd Workshop on Continuous Vector space Models and Their Compositionality*, 2015.

Sauleh Eetemadi and Kristina Toutanova. 2015. "Detecting Translation direction: A Cross-Domain Study". *Proceedings of NAACL Student Workshop*, 2015.

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.*

INVITED TALKS

"Joint Compositional Learning from Text and Knowledge Bases"
  Invited talk at AKBC, June 2016

"Beyond Left-to-Right: Multiple Decomposition Structures for SMT"
  University of Pennsylvania, April 2014
  Johns Hopkins University, April 2014

"Morphological Knowledge in Statistical Machine Translation"
  Keynote at Machine Translation Marathon, September 2013
  Keynote at Dagstuhl Seminar on Translation into morphologically Rich Languages, February 2014

"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

Journal Editor: Co-editor in chief, TACL Journal 2015-current

Executive board: North American Chapter of the Association for Computational Linguistics – Board Member (2010 – 2011)

Program Chair: ACL 2014 PC co-chair, CoNLL 2008 PC co-chair

Area Chair: NAACL 2009 (machine learning), EMNLP 2010 (semantics), EMNLP 2012 (syntax and parsing), NAACL 2016.

Thesis committee: external examiner for Yudong Liu, Simon Fraser University, thesis committee for ThuyLinh Nguyen, Carnegie Mellon University, Jason Naradowsky, University of Massachusetts Amherst, Kairit Sirts, University of Edinburgh, Sauleh Eetemadi, University of Missouri.

Journal committee: editorial board of Computational Linguistics (2012-2014), TACL action editor 2012-2015

Conference Reviewing: ACL, NAACL, EMNLP, ICML, NIPS, AAAI, COLING, CONLL, IJCAI.

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

LANGUAGES

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