

DAN KLEIN
CURRICULUM VITAE

(510) 643-0805
klein@cs.berkeley.edu

CURRENT

Assistant Professor, Computer Science Division, UC Berkeley *2004-Present*

EDUCATION

Stanford University *1999-2004*

Ph.D. in Computer Science

Advisor: Christopher D. Manning

Dissertation: The Unsupervised Learning of Natural Language Structure

Stanford University *1999-2000*

Master of Science in Computer Science, with Distinction

Oxford University, St. John's College *1998-1999*

Master of Studies in Linguistics, with Distinction

Cornell University, School of Arts and Sciences *1994-1998*

Bachelor of Arts, *Summa Cum Laude*

Majors in Computer Science, Linguistics, and Mathematics

HONORS AND AWARDS

ACM Grace Murray Hopper Award *2007*

Alfred P. Sloan Fellowship *2007*

National Science Foundation CAREER Award *2006*

Hellman New Faculty Fellowship *2006*

Microsoft Research New Faculty Fellowship *2005*

Microsoft Graduate Research Fellowship *2003*

British Marshall Scholarship *1998*

Stanford Graduate Fellowship *1998*

National Science Foundation Graduate Fellowship *1998*

Department of Defense Fellowship (declined) *1998*

Merrill Presidential Scholar, Cornell University *1998*

Beinecke Brothers Memorial Fellowship *1997*

Barry M. Goldwater Fellowship *1996*

Best Paper Awards:

NAACL-06 for "Prototype-Driven Sequence Models" *2006*

EMNLP-04 for "Max-Margin Parsing" *2004*

ACL-03 for "Accurate Unlexicalized Parsing" *2003*

RESEARCH INTERESTS

Natural Language Processing (Unsupervised Language Learning, Efficient Parsing, Machine Translation, Information Extraction and Retrieval), AI, Machine Learning

PUBLICATIONS

Thesis

Dan Klein, "The Unsupervised Learning of Natural Language Structure," Ph.D. thesis, Stanford University, 2005.

Refereed Publications

Dan Klein and Christopher D. Manning, "Parsing with Treebank Grammars: Empirical Bounds, Theoretical Models, and the Structure of the Penn Treebank," *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 2001.

Dan Klein and Christopher D. Manning, "Parsing and Hypergraphs," *Proceedings of the 7th International Workshop on Parsing Technologies*, 2001.

Dan Klein and Christopher D. Manning, "Distributional Phrase Structure Induction," *Proceedings of the Fifth Conference on Natural Language Learning*, 2001.

Dan Klein and Christopher D. Manning, "Natural Language Grammar Induction using a Constituent-Context Model," *Advances in Neural Information Processing Systems 14*, 2001.

Taher H. Haveliwala, Aristides Gionis, Dan Klein, and Piotr Indyk, "Evaluating Strategies for Similarity Search on the Web," *The Proceedings of the Eleventh International World Wide Web Conference*, 2002.

Dan Klein and Christopher D. Manning, "A Generative Constituent-Context Model for Improved Grammar Induction," *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, 2002.

Dan Klein and Christopher D. Manning, "Conditional Structure versus Conditional Estimation in NLP Models," *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2002.

Dan Klein and Christopher D. Manning, "Fast Exact Inference with a Factored Model for Natural Language Processing," *Advances in Neural Information Processing Systems 15*, 2002.

Sepandar D. Kamvar, Dan Klein, and Christopher D. Manning, "Interpreting and Extending Classical Agglomerative Clustering Algorithms using a Model-Based Approach," *Proceedings of the Nineteenth International Conference on Machine Learning*, 2002.

Dan Klein, Sepandar D. Kamvar, and Christopher D. Manning, "From Instance-level Constraints to Space-Level Constraints: Making the Most of Prior Knowledge in Data Clustering," *Proceedings of the Nineteenth International Conference on Machine Learning*, 2002.

Dan Klein and Christopher D. Manning, "Parsing and Hypergraphs," in Harry Bunt, John Carroll, and Giorgio Satta, eds., *New Developments in Parsing Technology*, Kluwer Academic Publishers, 2003.

Dan Klein, Joseph Smarr, Huy Nguyen, and Christopher D. Manning, "Named Entity Recognition with Character-Level Models," *Proceedings of the Seventh Conference on Natural Language Learning*, 2003.

Dan Klein and Christopher D. Manning, "A* Parsing: Fast Exact Viterbi Parse Selection," *Proceedings of the North American Chapter of the Association for Computational Linguistics 2003*, 2003

Sepandar D. Kamvar, Dan Klein, and Christopher D. Manning, "Spectral Learning," *Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 2003.

Dan Klein and Christopher D. Manning, "Factored A* Search for Models over Sequences and Trees," *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 2003.

Kristina Toutanova, Dan Klein, and Christopher D. Manning, "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," *Proceedings of the North American Chapter of the Association for Computational Linguistics*, 2003.

Dan Klein and Christopher D. Manning, "Accurate Unlexicalized Parsing," *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 2003. [Best Paper Award]

Dan Klein and Christopher D. Manning, "Corpus-Based Induction of Syntactic Structure: Models of Dependency and Constituency," *Proceedings of the 42nd Annual Meeting of the ACL*, 2004.

Dan Klein and Christopher D. Manning, "Natural Language Grammar Induction with a Generative Constituent-Context Model," *Pattern Recognition: Special Issue on Grammatical Inference*, Elsevier, April 2004.

Ben Taskar, Dan Klein, Michael Collins, Daphne Koller, and Christopher D. Manning, "Max-Margin Parsing," *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2004. [Best Paper Award]

Trond Grenager, Dan Klein, and Christopher D. Manning, "Accurate Unsupervised Learning of Field Structure Models for Information Extraction," *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 2005.

Yang Huang, Henry Lowe, Dan Klein, and Russell Cucina, "Improved Identification of Noun Phrases in Clinical Radiology Reports using a High Performance Statistical Natural Language Parser, Augmented with the UMLS Specialist Lexicon," *Journal of the American Medical Informatics Association*, 2005.

Ben Taskar, Simon Lacoste-Julien, and Dan Klein, "A Discriminative Matching Approach to Word Alignment", *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2005

Percy Liang, Ben Taskar, and Dan Klein, "Alignment by Agreement," *Proceedings of the North American Chapter of the Association for Computational Linguistics*, 2006.

Simon Lacoste-Julien, Ben Taskar, and Dan Klein, "Word Alignment via Quadratic Assignment," *Proceedings of the North American Chapter of the Association for Computational Linguistics*, 2006.

Aria Haghighi and Dan Klein, "Prototype-Driven Learning for Sequence Models," *Proceedings of the North American Chapter of the Association for Computational Linguistics*, 2006. [Best Student Paper Award]

Percy Liang, Alexandre Bouchard-Cote, Dan Klein, and Ben Taskar, "An End-to-End Discriminative Approach to Machine Translation," *Proceedings of the Association for Computational Linguistics*, 2006.

Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein, "Learning Accurate, Compact, and Interpretable Tree Annotation," *Proceedings of the Association for Computational Linguistics*, 2006.

Slav Petrov, Leon Barrett, and Dan Klein, "Non-Local Modeling with a Mixture of PCFGs," *Proceedings of the Conference on Natural Language Learning*, 2006.

Aria Haghighi and Dan Klein, "Prototype-Driven Grammar Induction," *Proceedings of the Association for Computational Linguistics*, 2006.

Alexandre Bouchard-Côté, Percy Liang, Thomas Griffiths, and Dan Klein, “A Probabilistic Approach to Diachronic Phonology,” *Proceedings of EMNLP-CoNLL*, 2007.

Percy Liang, Slav Petrov, Michael Jordan, and Dan Klein, “The Infinite PCFG using Hierarchical Dirichlet Processes,” *Proceedings of EMNLP-CoNLL*, 2007.

Slav Petrov, Adam Pauls, and Dan Klein, “Learning Structured Models for Phone Recognition,” *Proceedings of EMNLP-CoNLL*, 2007.

Aria Haghighi and Dan Klein, “Unsupervised Coreference Resolution in a Nonparametric Bayesian Model,” *Proceedings of ACL*, 2007.

John DeNero and Dan Klein, “Tailoring Word Alignments to Syntactic Machine Translation,” *Proceedings of ACL*, 2007.

Slav Petrov and Dan Klein, “Improved Inference for Unlexicalized Parsing,” *Proceedings of HLT-NAACL*, 2007.

Aria Haghighi, John DeNero, and Dan Klein, “Approximate Factoring for A* Search,” *Proceedings of HLT-NAACL*, 2007.

Aria Haghighi, John DeNero, and Dan Klein, “A* Search via Approximate Factoring,” *Proceedings of AAAI (Nectar Track)*, 2007.

Slav Petrov and Dan Klein, “Learning and Inference for Hierarchically Split PCFGs,” *Proceedings of AAAI (Nectar Track)*, 2007.

David Rosenberg, Dan Klein, and Ben Taskar, “Mixture-of-Parents Maximum Entropy Markov Models,” *Proceedings of UAI*, 2007.

Technical Reports and Workshop Papers

Taher H. Haveliwala, Aristides Gionis, Dan Klein, and Piotr Indyk, “Similarity Search on the Web: Evaluation and Scalability Considerations,” *Stanford Technical Report*, 2000.

Dan Klein and Christopher D. Manning, “An $O(n^3)$ Agenda-Based Chart Parser for Arbitrary Probabilistic Context-Free Grammars,” *Stanford Technical Report*, 2001.

Dan Klein, Kristina Toutanova, H. Tolga Ilhan, Sepandar D. Kamvar, and Christopher D. Manning, “Combining Heterogeneous Classifiers for Word-Sense Disambiguation,” *Word Sense Disambiguation Workshop at the 40th Annual Meeting of the Association for Computational Linguistics*, 2002.

Dan Klein, and Christopher D. Manning, “A Core-Tools Statistical NLP Course,” *Proceedings of the 2nd Workshop on Teaching NLP*, 2005.

John DeNero, Dan Gillick, James Zhang, and Dan Klein, “Why Generative Phrase Models Underperform Surface Heuristics,” *Workshop on Statistical Machine Translation at NAACL*, 2006.

Other Publications

Dan Klein, “Statistical Parsing,” in *Encyclopedia of Language and Linguistics, 2nd Edition*, Keith Brown, editor, Elsevier, 2005.

Dan Klein, “Review of *Data-oriented Parsing* edited by R. Bod *et al.*,” *Computational Linguistics*, 30:2, pp. 240-244, MIT Press, June 2004.

Dan Klein and Christopher D. Manning, “Maxent Models, Conditional Estimation, and Optimization, without the Magic,” tutorial presented at NAACL 2003 and ACL 2003.

Dan Klein and Ben Taskar, “SVMs and Structured Max-Margin Methods,” tutorial presented at ACL 2005.

PROFESSIONAL ACTIVITIES

Software / Tools: My older parsers are available at <http://nlp.stanford.edu/downloads/lex-parser.shtml> and have been used by dozens of researchers and projects. Several new software packages are at <http://nlp.cs.berkeley.edu/>.

Selected Invited Talks:

“Statistical Methods for Machine Translation,” JSM 2006

“Unsupervised Methods for Natural Language,” JHU Summer Workshop 2005

“NLP: Unsupervised, Adaptive, and Efficient Technologies,” BEARS 2005

“NLP: Unsupervised and Adaptive Methods,” ICSI Speech Group, 2005.

“Unsupervised Learning of Natural Language Structure,” Johns Hopkins University, 2004.

“Constrained Spectral Clustering,” Intel Research, NASA Ames Research, 2004.

“Unsupervised Learning for Natural Language” (Academic Job Talk) UC Berkeley, CMU, UW Seattle, Princeton, Brown, UCSD, Caltech, University of Chicago / TTI, Microsoft Research, 2004.

Tutorials:

“Maxent Models, Conditional Estimation, and Optimization, without the Magic.” ACL 2003, NAACL 2003, with Chris Manning

“SVMs and Structured Max-Margin Methods,” ACL 2005, with Ben Taskar

“Machine Learning for Natural Language Processing: New Developments and Challenges,” NIPS 2006

“Introduction to Classification: Likelihoods, Margins, Features, and Kernels,” NAACL 2007

“Bayesian Nonparametric Structured Models,” ACL 2007, with Percy Liang

Conference Reviewing: ACL, NAACL, EMNLP, ICML, NIPS, IJCAI.

Journal Reviewing: *Computational Linguistics*, *Machine Learning Journal*, *Journal of Natural Language Engineering*, *Journal of Machine Learning Research*, *Computers and the Humanities*, *Computer Speech and Language*, *PLoS Computational Biology*.

Manuscript Reviewing: CSLI Press.

Other External Service: NSF funding panel (2004), Co-chair of the Tenth Conference on Natural Language Learning (2006), Area Chair for NAACL 2007, Area Chair for NIPS 2007.

TEACHING EXPERIENCE

CS188: Introduction to Artificial Intelligence, UC Berkeley Sp06, Fa06, Fa07
Created a modified version of the upper-division undergraduate AI course, including preparing and delivering lectures, and the design of four applied programming projects.

CS294: Statistical Natural Language Processing, UC Berkeley Fa04, Fa05, Sp07
Created new graduate course, including preparing and delivering lectures and the creation of four hands-on programming projects.

LSA 322: Grammar Induction, *Linguistic Summer Institute* Su07

Created new graduate course, including preparing and delivering lectures and the creation of four hands-on programming projects.

CS161: Algorithms, *Stanford University* Su00

Prepared and delivered all course lectures to both a live and televised audience; designed course syllabus, designed homeworks, and exams; redesigned term project from scratch.