

## DANIEL M. BIKEL

1500 Locust Street, apt. #2608  
Philadelphia, PA 19102  
(215) 985-3158

<http://www.cis.upenn.edu/~dbikel/>

Institute for Research in Cognitive Science  
University of Pennsylvania  
3401 Walnut Street, Suite 400A  
Philadelphia, PA 19104-6228  
dbikel@cis.upenn.edu

### EDUCATION

**UNIVERSITY OF PENNSYLVANIA**, Philadelphia, PA.

**Ph.D. candidate** in the Dept. of Computer & Information Science (degree expected summer of 2004).

**M.S.E. in Computer Science** (May, 1999).

Advisor: Prof. Mitch Marcus. Research in statistical techniques for natural language processing, focusing on combining parsing and word sense disambiguation.

**HARVARD UNIVERSITY**, Cambridge, MA.

**Graduate student in Computer Science** (Department of Applied Sciences), through the Graduate School of Arts and Sciences Special Student Program.

**HARVARD UNIVERSITY**, Cambridge, MA.

**A.B. with honors in Classics—Greek & Latin** in June, 1993. Awarded Harvard College Scholarship for excellence in academics. Admitted as National Scholar.

**PHILLIPS EXETER ACADEMY**, Exeter, NH.

**Awarded Classics diploma with highest honors** in June, 1989. Winner of Teschemacher 4-year scholarship to Harvard College.

### PROFESSIONAL ACTIVITIES

**NATIONAL SCIENCE FOUNDATION panelist**, Human-Computer Interface/Natural Language Processing Panel (May, 2002).

**INSTITUTE FOR RESEARCH IN COGNITIVE SCIENCE (IRCS) Fellow** (2000 through 2003).

**COMPUTATIONAL LINGUISTICS JOURNAL** reviewer (2002).

**ASSOCIATION OF COMPUTATIONAL LINGUISTICS** conference program committee member (2001).

**EUROPEAN ASSOCIATION OF COMPUTATIONAL LINGUISTICS** conference program committee member (2003).

**EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING** conference program committee member (2003).

**HUMAN LANGUAGE TECHNOLOGY 2002** conference co-organizer.

### RESEARCH EXPERIENCE

**Doctoral research:** Computer and Information Science Department, University of Pennsylvania, 1997–present (research advisor: Prof. Mitch Marcus).

- Development and analysis of techniques to incorporate word sense information into statistical parsing models
- Development and analysis of techniques for using machine learning to augment heuristics used by statistical parsing algorithms
- State-of-the-art results in parsing Chinese

- Analysis of probabilistic dependencies and smoothing techniques in statistical parsing
- Full analysis of the widely-used Collins thesis parser; first ever duplication of parsing results of Collins' thesis parser
- Development of first scalable, portable statistical parsing engine that is extensible to new models and to new domains and human languages (<http://www.cis.upenn.edu/~dbikel/#stat-parser>)
- Development of techniques for bootstrapping the creation of treebanks (joint work with Prof. Tony Kroch of the Linguistics Dept. of the Univ. of Pennsylvania)

**Industrial research:** Speech and Language Group, BBN Technologies, Cambridge, MA

- Information extraction research
- Part-of-speech tagging research
- Statistical name-finding in natural language text

## PATENT

Nymble name-finding technology (developed while at BBN), U.S. Patent No. 6,052,682: "Method Of And Apparatus For Recognizing And Labeling Instances of Name Classes in Textual Environments" (granted April 18, 2000).

## WORK EXPERIENCE

**SPEECH AND LANGUAGE GROUP, BBN TECHNOLOGIES**, Cambridge, MA.

**Associate Scientist.** Research and development on components of group's information-extraction system, PLUM. Helped develop set of domain-knowledge editors to allow expert users to port PLUM from one domain to another ("NLU Shell"). Worked on improvement of stochastic part-of-speech taggers. Helped develop top-performing rule-based, finite-state pattern recognizer for finding names and other non-recursive entities. Built from scratch stochastic learning name-finder called Nymble, that out-performs all other learning name-finders and approaches performance of rule-based systems. (July, 1994 through July, 1997)

**DAN BIKEL RECORDING**, Wilton, CT.

**Self-employed entrepreneur**, culminating in the production of an *album of original recorded music* on cassette and Compact Disc (1000 units produced). Composed 13 pieces in rock, folk, jazz and classical genres, with full use of MIDI, multitrack and digital recording technology, including post-production involving extensive signal processing and digital editing. Designed graphics for CD and cassette inserts as well as CD silk-screen. Album launched in late spring of 1992, and was sold in several stores in Cambridge, MA. (Summers of 1989, 1990 and 1991.)

## TEACHING

**Teaching Assistant at the University of Pennsylvania**

- CSE110 Introduction to Programming (Fall, 1998).
- CSE350 Software Design/Engineering (Spring 1999).
- CSE100 Information Technology and Its Impact on Society (Fall, 1999).

**Independent Study Supervisor at the University of Pennsylvania** for two undergraduate students and co-supervisor for one Master's student.

## PUBLICATIONS

### JOURNAL ARTICLES

Daniel M. Bikel. 2003. Intricacies of Collins' Parsing Model. TR No. MS-CIS-03-11 (*accepted with revisions to the Computational Linguistics journal*).

Daniel M. Bikel, Richard Schwartz and Ralph M. Weischedel. 1999. An Algorithm that Learns What's in a Name. In the Machine Learning Journal Special Issue on Natural Language Learning.

### CONFERENCE ARTICLES

David Chiang and Daniel M. Bikel. 2002. Recovering Latent Information in Treebanks. In the proceedings of COLING 2002.

Daniel M. Bikel. 2002. Design of a Multi-lingual, Parallel-processing Statistical Parsing Engine. In the proceedings of HLT 2002.

Daniel M. Bikel. 2000. A Statistical Model for Parsing and Word-Sense Disambiguation. In the proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora.

Daniel M. Bikel, Scott Miller, Richard Schwartz and Ralph Weischedel. 1997. Nymble: a High-Performance Learning Name-finder. In the proceedings of the Conference on Applied Natural Language Processing, '97.

### WORKSHOP ARTICLES

Daniel M. Bikel and David Chiang. 2000. Two Statistical Parsing Models Applied to the Chinese Treebank. In the proceedings of the Second Chinese Language Processing Workshop.

## REFERENCES

Professor Mitch Marcus (Ph.D. advisor)  
Dept. of Computer & Information Science  
University of Pennsylvania  
503 Levine Hall, 3330 Walnut Street  
Philadelphia, PA 19104-6389  
email: mitch@cis.upenn.edu  
tel.: (215) 898-2538  
fax: (215) 898-0587

Professor Fernando Pereira  
Dept. of Computer & Information Science  
University of Pennsylvania  
305 Levine Hall, 3330 Walnut Street  
Philadelphia, PA 19104-6389  
email: pereira@cis.upenn.edu  
tel. (215) 573-5041  
fax: (215) 898-0587

Professor Aravind Joshi  
Institute for Research in Cognitive Science,  
Suite 400A  
3401 Walnut Street  
Philadelphia, PA 19104-6228  
email: joshi@cis.upenn.edu  
tel.: (215) 898-0359  
fax: (215) 573-9247

Professor Martha Palmer  
Dept. of Computer & Information Science  
University of Pennsylvania  
504 Levine Hall, 3330 Walnut Street,  
Philadelphia, PA 19104-638  
email: mpalmer@cis.upenn.edu  
tel.: (215) 898-9513  
fax: (215) 898-0587

Professor Mark Johnson  
Box 1978, Brown University  
Providence, Rhode Island 02912  
email: mj@cs.brown.edu  
tel.: (401) 863-1670

Professor Anthony Kroch  
Dept. of Linguistics  
University of Pennsylvania  
Philadelphia, PA 19104-6305  
email: kroch@change.ling.upenn.edu  
tel.: (215) 573-4794  
fax: (215) 573-2091

## **ADDITIONAL INFORMATION**

U.S. Citizen.

Languages: native English speaker, reading ability in Ancient Greek and Latin.

## **ADDITIONAL ACTIVITIES**

Road cyclist. Joined the **PENN CYCLING TEAM** in 2001, and competed in over 30 races in the 2002 season.

Argentine tango.

Guitarist, keyboardist, singer/songwriter. Music arranging and recording.

**HARVARD GLEE CLUB** member and soloist for four years, performing and soloing in choral concerts around the world.

Extensive travel throughout the Americas, the British Isles, Europe, the Middle East, Africa and parts of East Asia.