

Curriculum Vitae

PAUL C. DAVIS

304 E. Grove Street
Arlington Heights, Illinois 60005
phone: (847) 818-9553 email: paulcdavis@gmail.com

AREAS OF SPECIALIZATION

Natural Language Processing; Machine Translation, with a focus on statistical MT; Information Extraction; Computational Pragmatics, including models of discourse and question answering; Machine Learning.

EDUCATION

Ph.D., Linguistics, Ohio State University, June 2002. Specializing in Computational Linguistics.
Dissertation: *Stone Soup Translation: The Linked Automata Model*.

Dissertation Committee: Chris Brew, Advisor (Ohio State, Linguistics); Detmar Meurers (Ohio State, Linguistics); Erhard Hinrichs (Universität Tübingen, Linguistics); and Robert Kasper (Mount Vernon Nazarene College, Computer Science).

M.A., Linguistics, University of Wisconsin-Madison, 1996.

B.S., Business Economics, Miami University, 1988, Cum Laude. Minor in International Business.
Attended Miami University European Center in Luxembourg, 1986-1987.

Relevant coursework:

Phonetics, Phonology, Historical Linguistics, Psycholinguistics, Syntax, Semantics, and Pragmatics.

Computational Linguistics seminars in Parsing, Computational Semantics, Constraint-Based Grammar Implementation, Finite-State NLP, Statistical NLP, Statistical Modeling.

Computer Science: Computer Architecture, Operating Systems, and Programming Languages (Ohio State University); Natural Language Processing, Data Structures and Algorithms, Artificial Intelligence, and Seminar in Machine Learning (University of Wisconsin-Madison).

Mathematics: Advanced Calculus, Algebraic Linguistics, and Category Theory.

PROFESSIONAL EXPERIENCE

2004–Present **Principal Staff Research Engineer**

NLP GROUP, HUMAN INTERFACE LABORATORY, MOTOROLA, SCHAUMBURG, ILLINOIS.

Led research focus area in statistical and symbolic natural language processing. Designed and implemented distributed, multiple-user interaction system supporting multiple modalities such as GUI, voice, and text. Directed development of system for automatic translation of natural language text.

2002–2004 **Senior Staff Research Engineer**

NLP GROUP, HUMAN INTERFACE LABORATORY, MOTOROLA, SCHAUMBURG, ILLINOIS.

Specialized in the design, feature enhancements, optimization, and porting of multilingual dialogue systems, parsers, and related software in a variety of computing environments and device architectures.

1996–2002

Research Assistant

DEPARTMENT OF LINGUISTICS, OHIO STATE UNIVERSITY, COLUMBUS, OHIO.

Computational linguistics projects including the following major projects:

- 2001-2002** Development of a finite-state probabilistic machine translation (MT) system, sponsored by a Motorola University Partnerships in Research Grant (with Chris Brew).

A new finite-state device involving the linking of automata was developed to more directly capture word-order differences than can be accomplished via transducers. The approach, known as the linked automata model, probabilistically links transition sequences of source and target language automata via an alignment table. The table represents an alignment model, and the two automata represent source and target language models, respectively.

The project involved the design of the linked automata approach, including algorithms for training, translation, and generalization; the development of probabilistic versions of the models; methods for the extraction of partial results; as well as the design of a specialized finite-state library. Techniques for word alignment and word alignment evaluation were also created, along with a graphical word alignment tool.

The project culminated in the development of the linked automata MT system. In addition to new flexibility in terms of word-ordering differences between two languages, this MT method allows for the representation of all possible alignments of sequences of words, including those alignments that are discontinuous. The model thus yields increased coverage within the efficiency of a finite-state framework.

- 1998-2001** Development and implementation of a natural language dialogue system, sponsored by a Motorola University Partnerships in Research Grant (with Robert Kasper and Craig Roberts).

The project focused on the computational implementation of a theory of pragmatics based on question-answering. Research centered on the appropriate design of presuppositional operators in the environment of a human/machine hotel reservation booking system, and included the integration of the presuppositional operators with a syntactic parser (LKB), and a specially designed domain ontology, using the Loom knowledge representation framework. Research also included the development of a planning module and a simple template-based generator. Programming was done mainly in C++. Cross-platform communication interfaces between various components running under Unix and Windows NT were also developed in both C++ and Lisp. The project resulted in the creation of a demonstration system presented to Motorola management in September 2000.

- 1997-1998** Computational analysis of political texts, in conjunction with the Mershon Center for International Security and Public Policy (with Robert Kasper).

This project involved information extraction from texts of speeches by key political figures in the Middle East. Processing included parsing the texts, resolving contextual information, and extracting relevant information. Implementation involved the use of Perl scripts for text preparation and selection, and Lisp programs for contextual resolution, as well as development of a specialized lexicon and integration of the LKB parser, and the design of a set of semantic representations for input of the extracted information into a Prolog database created for the project.

1997 Parsing of languages with flexible constituent ordering (with Robert Kasper).

This project involved designing a parser able to efficiently handle (semi-) free word-order languages such as German in a constraint-based environment. The system successfully implemented a German grammar fragment in HPSG which could efficiently process all allowable word orders.

1999, 2000 (Summer) Software Programming Intern

NLP GROUP, HUMAN INTERFACE LABORATORY, MOTOROLA, SCHAUMBURG, ILLINOIS.

Ported a natural language dialogue system from Lisp to C⁺⁺. Continued the design, development, and integration of modules of the dialogue system, and created a demonstration system.

1997 (Summer) Research Assistant

LINGUISTICS DEPARTMENT, EBERHARD KARLS UNIVERSITÄT, TÜBINGEN, GERMANY.

Developed strategies and software tools for syntactic annotation of English Treebank for the Verbmobil project, using both phrasal and functional annotations (i.e., both tree nodes and edges were labeled). Tools developed were implemented in Perl and included an automated lexicon and summarization tool to assist in viewing previous annotations. Annotated several hundred sentences.

1994-1996 Editorial Assistant

ADVANCING THE CONSUMER INTEREST, MADISON, WISCONSIN.

Solicited, edited, and prepared manuscripts for review and publication in an academic journal focusing on consumer law, policy, and research.

1992-1993 Reforestation Agent

DEVELOPMENT CORPORATION OF CHUQUISACA / U.S. PEACE CORPS, CULPINA, BOLIVIA.

Initiated rural reforestation project; secured funding and external agency assistance; administered coursework in agribusiness practices.

1988-1992 Assistant Office Manager / Paralegal

BLACKBURN & MICHELMAN, P.C., PHILADELPHIA, PENNSYLVANIA.

Managed billing and computing resources; analyzed damages, managed discovery documents, and supervised legal assistants.

PUBLICATIONS

All Links are not the Same: Evaluating Word Alignments for Statistical Machine Translation. Paul C. Davis, Zhuli Xie, and Kevin Small. *Machine Translation Summit XI Proceedings (MT Summit XI)*, pp. 119-126. Copenhagen, Denmark. September, 2007.

Recognizing and Organizing Opinions Expressed in the World Press. Janyce Wiebe, Eric Breck, Chris Buckley, Claire Cardie, Paul Davis, Bruce Fraser, Diane Litman, David Pierce, Ellen Riloff, Theresa Wilson, David Day, and Mark Maybury. *Working Notes of the 2003 AAIL Spring Symposium on New Directions in Question Answering*. Stanford, CA. March, 2003.

NRRC Summer Workshop on Multiple-Perspective Question Answering: Final Report. Janyce Wiebe, Eric Breck, Chris Buckley, Claire Cardie, Paul Davis, Bruce Fraser, Diane Litman, David Pierce, Ellen Riloff, and Theresa Wilson. 2002.

Stone Soup Translation: The Linked Automata Model. Paul C. Davis. Ph.D. Dissertation in Linguistics. Ohio State University. 2002.

Stone Soup Translation. Paul C. Davis and Chris Brew. *Proceedings of the 9th Conference on Theoretical and Methodological Issues in Machine Translation (TMI-2002)*, pp. 31-41. Keihanna, Japan. March, 2002.

Presupposition Resolution with Discourse Information Structures. Paul C. Davis. *Ohio State University Working Papers in Linguistics, No. 54*, pp. 25-58. 2000.

An Integrated Approach to Reference and Presupposition Resolution. Robert T. Kasper, Paul C. Davis, and Craig Roberts. *Proceedings of the Workshop: The Relation of Discourse/Dialogue Structure and Reference, 37th Annual Meeting of the Association for Computational Linguistics*, pp. 1-10. College Park, Maryland. June, 1999.

Know When to Hold 'Em: Shuffling Deterministically in a Parser for Nonconcatenative Grammars. Robert T. Kasper, Mike Calcagno, and Paul C. Davis. *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and the 17th International Conference on Computational Linguistics (COLING-ACL '98)*, pp. 663-669. Montreal, Canada. August, 1998.

PRESENTATIONS

“Stone Soup Translation.” 9th Conference on Theoretical and Methodological Issues in Machine Translation (TMI-2002). Keihanna, Japan. March 13, 2002.

“Naïve Finite-State Translation.” Summer 2001 GRA in Cognitive Science Presentation, Center for Cognitive Science, Ohio State University, Columbus, Ohio. January 25, 2002.

“Stone Soup Translation.” Second Pre-Generals presentation, Department of Linguistics, Ohio State University, Columbus, Ohio. April 11, 2001.

“Modeling a Natural Language Dialogue System.” Invited presentation at Motorola Scientific Advisory Board Association meeting, Atlanta, Georgia. October 21, 2000.

“Presupposition Resolution with Discourse Information Structures.” First Pre-Generals presentation, Department of Linguistics, Ohio State University, Columbus, Ohio. April 24, 2000.

“Adverbial Modifiers: Linear Position and Semantic Function,” Linguistics Colloquium Fest, Ohio State University, Columbus, Ohio. June 6, 1998.

AWARDS

Selected for the Northeast Regional Research Center (NRRC) Workshop on Multiperspective Question Answering, 2002.

Recipient of four-year Motorola University Partnerships in Research Grant, 1998-2002.

Summer GRA in Cognitive Science, Ohio State University, 2001.

SERVICE

Member, Program Committee for Student Session of ACL 2000.

Member, Search Committee for Senior Computational Linguist, Department of Linguistics, Ohio State University, 1998.

Laboratory and Computing Committee, Department of Linguistics, Ohio State University, 1996-1997.

NATURAL LANGUAGES

Spanish, German, French, Quechua

PROGRAMMING LANGUAGES / ENVIRONMENTS / WEB TECHNOLOGIES

C++, Perl, Lisp, Prolog / Windows, Unix, Linux (Pentium, ARM), Symbian, Telematics environments / JavaScript, Ajax, XSLT, XPath, XML Schema, XForms

REFERENCES

Professor Chris Brew
Department of Linguistics
Ohio State University
1712 Neil Ave.
Columbus, Ohio 43210
email: cbrew@ling.osu.edu
phone: (614) 292-5420 fax: (614) 292-8833

Professor Detmar Meurers
Department of Linguistics
Ohio State University
1712 Neil Ave.
Columbus, Ohio 43210
email: dm@ling.osu.edu
phone: (614) 292-0461 fax: (614) 292-8833

Professor Erhard Hinrichs
Seminar für Sprachwissenschaft
Universität Tübingen
Wilhelmstraße 113
D-72074 Tübingen Germany
email: eh@sfs.nphil.uni-tuebingen.de
phone: (49) 7071-29-74279 fax: (49) 7071-55-1335

Professor Robert Kasper
Computer Science Department
Mount Vernon Nazarene College
800 Martinsburg Road
Mount Vernon, Ohio 43050
email: rkasper@mvnc.edu
phone: (740) 397-9000 ext. 4271 fax: (740) 397-2769