
KARIN M. VERSPOOR

2833 Don Quixote
Santa Fe, NM 87505

(505) 473-1715 (h) / (505) 667-5086 (o)

<http://public.lanl.gov/verspoor>
karin.verspoor@gmail.com

RESEARCH STATEMENT

I am a computational linguistics researcher with particular expertise in methods for representing and manipulating semantic information, and a research agenda oriented around developing hybrid methodologies combining statistical and quantitative analysis with symbolic, logical, and qualitative techniques to support robust yet semantically-enabled data analysis. My research is aimed at exploitation of information semantics to support information extraction from text and predictive analytics. I have specific interest and experience in the area of biomedical text processing, drawing on knowledge resources like ontologies to enable deep semantic extraction in areas such as protein-protein interaction extraction and protein function prediction.

EDUCATION

UNIVERSITY OF EDINBURGH – *Edinburgh, Scotland, UK* **1997**
PhD, Cognitive Science and Natural Language Thesis: “Contextually-Dependent Lexical Semantics”

UNIVERSITY OF EDINBURGH – *Edinburgh, Scotland, UK* **1994**
MSc, Cognitive Science and Natural Language Thesis: “A Cognitively-Relevant Lexical Semantics”
Awarded with Distinction

RICE UNIVERSITY – *Houston, TX* **1993**
BA, Computer Science and Cognitive Sciences Thesis: “What are the Characteristics of Emotional Metaphors?”
Graduated *Summa Cum Laude*

PROFESSIONAL EXPERIENCE

LOS ALAMOS NATIONAL LABORATORY – *Los Alamos, New Mexico* **2003 – present**
Team Leader, Knowledge and Language Science team 2007 - present
Technical Staff Member, Knowledge and Information Systems Science team 2003 - 2007
Performed scientific research in computational methods for the extraction, representation, organization, synthesis, discovery and retrieval of knowledge in databases and information systems. Primary applications in bioinformatics and homeland security.

APPLIED SEMANTICS, INC. – *Los Angeles, California* **2001 - 2002**
Senior Computational Linguist
Led core technology development of an enterprise suite offering semantic meta-tagging, summarization and categorization of documents. Developed algorithms for word sense disambiguation and “page sensing” through incorporation of NLP techniques such as reference resolution and shallow parsing. Augmented lexicon and ontology specificity through implementation of statistical techniques.

WEBMIND, INC. – *New York, New York* **1998 - 2001**
Director, Natural Language Engineering
Designed and implemented the natural language module in the Webmind Artificial Intelligence system, a module aimed at deep semantic understanding of texts and intelligent, conversational query answering. Directed a team of nine scientists and programmers, leading conceptual and implementation-level design.

MICROSOFT RESEARCH INSTITUTE – *Sydney, Australia* **1997 - 1998**
Research Fellow
Conducted research in multilingual natural language generation, and served as Project Coordinator of the Dynamic Document Delivery project, generating multilingual textual descriptions of museum objects from a database.

KARIN M. VERSPOOR

TEACHING EXPERIENCE

RICE UNIVERSITY – *Houston, TX*

Fall 1990 & Spring 1992, Teaching Assistant

Spring 1993, Teaching Assistant

Data Structures and C++ Programming

Artificial Intelligence

UNIVERSITY OF EDINBURGH, DEPT. OF ARTIFICIAL INTELLIGENCE – *Edinburgh, UK*

Fall 1994, Spring 1995, Fall 1995, Tutor/Marker

Introduction to Natural Language Processing

UNIVERSITY OF EDINBURGH, CENTRE FOR COGNITIVE SCIENCE – *Edinburgh, UK*

Fall 1994: Demonstrator

Fall 1994: Demonstrator

Fall 1995: Head Tutor & Demonstrator

Fall 1995: Head Tutor & Demonstrator

Spring 1996: Head tutor

Logic Programming

Computational Linguistics

Logic Programming

Computational Linguistics

Computational Semantics

MICROSOFT RESEARCH INSTITUTE – *Sydney, Australia*

1997-1998: Co-teaching with Robert Dale

Introduction to Natural Language Processing

RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING CONFERENCE – *Borovets, Bulgaria*

Sept. 2007: TUTORIAL developed with Kevin B. Cohen Natural Language Processing and the Biomedical Domain

GRANTS AND CONTRACTS AWARDED

- Los Alamos National Laboratory Exploratory Research Grant (PI), 2006-2008 \$775,000
Automated Induction of Templates for Extracting Information from Text
- Defense Threat Reduction Agency, 2006-present
Technology Innovation Consortium
- Department of Homeland Security, 2006-2007
Motivation and Intent program
- Department of Homeland Security, 2005-2007
Analysis, Dissemination, Visualization, Insight and Semantic Enhancement program
- Los Alamos National Laboratory-New Mexico State University collaboration program, 2005-2006
Modeling and Analyzing Social Behavior: Applications toward National Security
- Cooperative Research and Development Agreement, Proctor & Gamble Corporation, 2003-2004
Cellular Pathway Discovery through Natural Language Knowledge Systems
- Macquarie University Research Grant (Principal Investigator), 1997 \$7,000
Building multilingual resources for Natural Language Generation
- UK Engineering and Physical Sciences Research Council Research Studentship for PhD study, 1994-1997
- Doctor Catherine van Tussenbroek grant to fund PhD research visits to the University of Utrecht and Brandeis University, 1995-1996

PROFESSIONAL SERVICE

- Panelist, National Science Foundation, CAREER award review panel, November 2007
- Panelist, kick-off workshop for National Science Foundation “Cyber-enabled Discovery and Innovation” program, October 2007
- Member, Organizing Committee UCLA Institute for Pure and Applied Mathematics (IPAM) “Mathematics of Knowledge and Search Engines” workshop series, Fall 2007
- Chair, UCLA IPAM Workshop on “Dynamic Searches and Knowledge Building”, October 2007
- Member, Organizing Committee UCLA IPAM “Social Data Mining and Knowledge Building” workshop, November 2007

KARIN M. VERSPOOR

- Member, OASIS-Open “Unstructured Information Management Architecture” standards committee, October 2006-present
- Chair, BioNLP’06. Linking Natural Language Processing and Biology, Workshop at NAACL 2006
- Adjunct Professor, Communication Studies Department, New Mexico State University, 2005-2006
- Program Committees and reviewing for numerous conferences, workshops, and journals
 - Journals: Genome Biology, Journal of Biomedical Discovery, PLoS Computational Biology, Journal of the American Medical Informatics Association, SKY Journal of Linguistics, Journal of User-Modeling and User-Adapted Interaction
 - Meetings: Pacific Symposium of Biocomputing, International Conference on Intelligent Text Processing and Computational Linguistics, BioCreative meetings, Intelligent Systems for Molecular Biology (Text Mining session; BioLink/BioOntologies workshops), European Conference on Computational Biology, Recent Advances in Natural Language Processing, Association for Computational Linguistics BioNLP Workshop, International Workshop on Knowledge Discovery in Life Science Literature, IEEE Systems Man and Cybernetics, Annual Meeting of the Cognitive Science Society, Perspectives of Systems Informatics, Artificial General Intelligence conference

INSTITUTIONAL SERVICE

- Los Alamos National Laboratory, Member, Integration of Heterogeneous Information Sources working group, LANL Grand Challenges committee, 2007-
- Los Alamos National Laboratory, Exploratory Research Grants Computer Science review committee, 2006
- Los Alamos National Laboratory, Women’s Diversity Working Group, Co-Chair 2006-
- Los Alamos National Laboratory, Diversity/Affirmative Action Board, Secretary 2005-2007

AWARDS

- 2007. Los Alamos National Laboratory, Computer and Computational Sciences Division, “Spot” award for excellence in promoting diversity at LANL.
- 1992. Rice Engineering Alumni Award for Top Junior Engineering Student.
- 1992. Induction to *Phi Beta Kappa* and *Tau Beta Pi* honor societies.

PROFESSIONAL AFFILIATIONS

- Association for Computational Linguistics, Member 1995-
- Linguistics Society of America, Member 1995-2006
- International Society for Computational Biology, Member 1994-2006

TECHNICAL EXPERTISE

Computational linguistics research experience:

- Information extraction, ontology-based methods for semantic analysis, lexical semantics, word sense disambiguation, coreference resolution and natural language generation.
- Language modeling, latent semantic analysis and terminology induction.
- Extensive background in theoretical linguistics and formal semantics.

Extensive software engineering and data management experience, including:

- Programming languages: Java, Perl, C++, Lisp, UNIX scripting tools, some Python.
- Database systems: Using ORACLE, DB2, SQL Server for data storage and analysis.

KARIN M. VERSPOOR

BIBLIOGRAPHY

JOURNAL PUBLICATIONS

1. **Verspoor, Karin.** (2005). Towards a semantic lexicon for biological language processing. *Comparative and Functional Genomics*, vol. 6, issue 1-2, p. 61-66. DOI: 10.1002/cfg.451.
2. **Verspoor, K.,** Cohn, J., Joslyn, C., Mniszewski, S., Rechtsteiner, A., Rocha, L.M., Simas, T. (2005). Protein Annotation as Term Categorization in the Gene Ontology using Word Proximity Networks. *BMC Bioinformatics* 2005, vol. 6 (suppl 1).
3. **Verspoor, K.,** Cohn, J., Mniszewski, S., and Joslyn, C. (2006). A Categorization Approach to Automated Ontological Function Annotation. *Protein Science*, v.15, pp.1544-1549.

EDITED COLLECTIONS

4. **Verspoor, K.,** Cohen, K.B., Goertzel, B. and Mani, I (eds.). (2006) BioNLP'06. Linking Natural Language Processing and Biology: Towards Deeper Biological Literature Analysis. Proceedings of the HLT-NAACL BioNLP Workshop.

PEER REVIEWED PUBLICATIONS

5. **Verspoor, C.M.** (1996a) A Perspective on PPs. *Edinburgh Working Papers in Cognitive Science*, Vol. 12: Studies in Head-Driven Phrase Structure Grammar, Claire Grover and Enric Vallduvi, eds., pp. 229-271.
6. **Verspoor, C.M.** (1996b) Lexical Limits on the Influence of Context. In proceedings Eighteenth Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates Publishers, pp. 116-120.
7. **Verspoor, C.M.** (1997) Conventionality-Governed Logical Metonymy. In proceedings 2nd International Workshop on Computational Semantics, H. Bunt, L. Kievit, R. Muskens and M. Verlinden, eds., pp. 300-312.
8. **Verspoor, C.M.** (1998) The author explores the issue onto the agenda: Predictivity vs. Stipulativity in the Lexicon In proceedings Pacific Asia Conference on Language, Information, and Computation, pp. 152-162.
9. **Verspoor, C.,** Dale, R., Green, S., Milosavljevic, M., Paris, C., and Williams, S. (1998) Intelligent Agents for Information Presentation: Dynamic Description of Knowledge Base Objects. In proceedings International Workshop on Intelligent Agents on the Internet and Web, Mexico City, Mexico, pp. 75-86.
10. Wan, S. and **Verspoor, C.** (1998), Automatic English-Chinese name transliteration for development of multilingual resources. In proceedings of COLING-ACL'98, Montreal, Canada.
11. Dale, R., Green, S., Milosavljevic, M., Paris, C., **Verspoor, C.** and Williams, S. (1998). Using Natural Language Generation Techniques to Produce Virtual Documents, In Proceedings of the Third Australian Document Computing Symposium (ADCS'98), Sydney, Australia.
12. Dale, R., Green, S., Milosavljevic, M., Paris, C., **Verspoor, C.,** Williams, S. (1998). The Realities of Generating Natural Language from Databases. In Proceedings 11th Australian Joint Conference on Artificial Intelligence, July 13-17, Brisbane, Australia.
13. Dale, R., Green, S., Milosavljevic, M., Paris, C., **Verspoor, C.,** Williams, S. (1998). Dynamic Document Delivery: Generating Natural Language Texts on Demand. In Proceedings of the 9th Conference and Workshop on Database and Expert Systems Applications (DEXA'98), August 24-28, Vienna, Austria.
14. **Verspoor, C.,** C. Joslyn and G. Papcun (2003). Interactions Between the Gene Ontology and a Domain Corpus for a Biological Natural Language Processing Application. In proceedings Sixth Annual Bio-Ontologies Meeting.
15. **Verspoor, C.,** C. Joslyn and G. Papcun (2003). The Gene Ontology as a Source of Lexical Semantic Knowledge for a Biological Natural Language Processing Application. In proceedings SIGIR'03 Workshop on Text Analysis and Search for Bioinformatics.
16. Joslyn, C., Cohn, J., **Verspoor, K.,** Mniszewski, S. (2005). Automating Ontological Function Annotation: Towards a Common Methodological Framework. In proceedings, BioOntologies SIG at ISMB 2005.
17. **Verspoor, K.,** Cohn, J., Mniszewski, S., Joslyn, C. (2005). POSOLE: Automated Ontological Annotation for Function Prediction. In participant notebook, Automated Function Prediction SIG at ISMB 2005.

KARIN M. VERSPOOR

18. Maguitman, A., Rechtsteiner, A., **Verspoor, K.**, Strauss, C.E., Rocha, L. (2006) Large-Scale Testing Of Bibliome Informatics Using Pfam Protein Families. Pacific Symposium of Biocomputing 11:76-87
19. Joslyn, C., Gessler, D.D.G., Schmidt, S.E., **Verspoor, K.M.** (2006) Distributed Representations of Bio-Ontologies for Semantic Web Services. In proceedings Joint BioLink and BioOntologies SIG of ISMB 2006.
20. **Verspoor, K.**, Sanfilippo, A., Elmore, M., MacKerrow, E. (2006). Deploying Natural Language Processing for Social Science Analysis. In proceedings Chicago Colloquium on Digital Humanities and Computer Science.
21. Abi-Haidar, A., Kaur, J., Maguitman, A., Radivojac, P., Retchsteiner, A., **Verspoor, K.**, Wang, Z., Rocha, L.M. (2007). Uncovering Protein-Protein Interactions in the Bibliome, in proceedings BioCreative II Workshop.
22. Joslyn, C., **Verspoor, K.**, Gessler, D. (2007). Knowledge Integration in Open Worlds: the Mathematics of Hierarchical Structure. In Proceedings of the First IEEE International Conference on Semantic Computing (ICSC'07).

OTHER PUBLICATIONS

23. Narasimhan, B., Di Tomaso, V. and **Verspoor, C.M.** (1996). Unaccusative or Unergative? Verbs of Manner of Motion. Quaderni del laboratorio di linguistica, 10, Scuola Normale Superiore of Pisa.
24. **Verspoor, C.M.**, G.J. Papcun and K. Sentz (2003). A Theoretical Motivation for Patterns in Information Extraction. Los Alamos Unlimited Release report 03-1504.
25. **Verspoor, K.** (2003) Text Mining for Bioinformatics. Poster at the Rocky Mountain Regional Bioinformatics Meeting.
26. **Verspoor, K.**, T. Simas, C. Joslyn, A. Rechtsteiner, S. Mniszewski, J. Cohn, L. Rocha, A. Fulmer (2004). Protein Annotation as Term Categorization in the Gene Ontology. Workshop paper, BioCreAtIvE meeting.
27. **Verspoor, K.**, C. Joslyn and G. Papcun (2004). A Lexical Semantic Network Induced from the Gene Ontology. Los Alamos Unclassified Report LAUR 04-0353.
28. **Verspoor, K.**, J. Cohn, S. Mniszewski, and C. Joslyn (2004). Nearest Neighbor Categorization for Function Prediction. In Proceedings of the 6th Community-wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction, Gaeta, Italy. December 4-8.
29. Cohn, J., **Verspoor, K.**, Mniszewski, S., and Joslyn, C. (2004). Predicting Protein Function Using Nearest Neighbor Categorization, Poster at 2nd Annual Rocky Mountain Regional Bioinformatics Meeting, Dec. 2004.
30. **Verspoor, K.M.**, Joslyn, C.A., Ambrosiano, J.A., Backer, A., Bodenreider, O., Hirschman, L., Karp, P., Kelly, H., Loranger, S., Musen, M., Sriram, R., and Wroe, C. (2005) Knowledge Integration for Biothreat Response, Los Alamos Unlimited Release report 05-0907
31. Joslyn, C., **Verspoor, K.**, Mniszewski, and Cohn, J., January 2005, Automated Protein Function Annotation Via Markov Adjusted Ontology Categorization, Los Alamos Unlimited Release report 05-0639.
32. **Verspoor, K.**, Cohn, J., Mniszewski, S., and Joslyn, C. (2005). Nearest Neighbor Categorization for CASP Function Prediction, poster presented at Intelligent Systems for Molecular Biology (ISMB 2005).
33. Joslyn, C., Mniszewski, S., **Verspoor, K.**, and Cohn, J. (2005). Improved Order Theoretical Techniques for GO Functional Annotation, poster at the Intelligent Systems for Molecular Biology (ISMB 2005).
34. Joslyn, C., Cohn, J., Mniszewski, S., **Verspoor, K.** (2005). Automating Ontological Function Annotation: Towards a Common Methodological Framework, poster at the BioOntologies SIG of the Intelligent Systems for Molecular Biology (ISMB 2005).
35. Joslyn, C., **Verspoor, K.**, Rodriguez, M. (2006). Representing the M&I Semantic Space: Concept Description for Lexical Management in Support of Ontology Development. Los Alamos Unlimited Release report 06-1060.
36. Gessler, D., Joslyn, C., Noy, N., Musen, M., **Verspoor, K.**, Bueno, R., Schmidt, S., Koester, B., Velculescu, V., Sjoblom, T. (2007). Visual Analytics and Ontologies for Discovering the Genetic Basis of Cancer. Los Alamos Unlimited Release report 07-1170.
37. Lin, S.-d. and **Verspoor, K.** (2007). A Semantics-Enhanced Language Model for Unsupervised Word Sense Disambiguation. Los Alamos Unlimited Release report 07-0618.

KARIN M. VERSPOOR

INVITED PRESENTATIONS (selected)

- *Statistics, Semantics, and Structure: examples of hybrid approaches to text processing problems*, invited presentations at Pacific Northwest National Laboratory and Thomson Research, October-November 2007
- *Exploiting contextual similarity for semantic prediction in the Biology domain*, invited workshop presentation, UCLA IPAM workshop on “Dynamic Searches and Knowledge Building”, October 2007
- *Mining Knowledge to Build Social Networks*, invited workshop presentation UCLA IPAM “Social Data Mining and Knowledge Building” workshop, November 2007
- *Semantics, Text, and the cure for Cancer*, Keynote speech, Recent Advances in Natural Language Processing conference, September 2007 (also given at MITRE Corp., Aug 2007)
- *Text Mining for Bioinformatics: Two projects at LANL*, December 2003. Presentation at MITRE Washington.
- *A Distributed Architecture for Artificial Intelligence*. Presented at the Association of Women in Science Central New Jersey Chapter meeting, June 2000.
- *The art of being a successful technical manager*. Presented at the Association for Women In Computing New York Chapter meeting, October 1999.
- *A Lexical Semantic Account of Prepositional Phrase Modification*. Presented at the 20th annual meeting of the DGfS (German Linguistics Society); Workshop on the Syntax of Adverbials, February 1998.
- *Phrasal constructs, the lexicon, and multilingual generation*. Presented at the Hong Kong University of Science and Technology Computer Science department, 17 February 1998. Also presented at Logos Corporation (Mount Arlington, NJ) and the Columbia University Computer Science department.