

---

<b>William D. Lewis</b>	Department of Linguistics University of Washington Box 354340 Seattle, WA 98195-4340	Tel: (206) 616-5728 Fax: (206) 685-7978 Email: wlewis2 AT u. washington. edu Web: faculty.washington.edu/wlewis2
-------------------------	---	---

---

### **Research Areas**

Corpus Linguistics, especially Web as Corpus  
Computational Linguistics (Automated Knowledge Discovery, Knowledge Representation,  
Ontologies, Semantic Networks, Web Mining, Corpus Building)  
Morphosyntax

### **Education**

**Ph.D. – Linguistics** (2002)

**Minor – Cognitive Science**

University of Arizona, Tucson, AZ

Dissertation: “A Distributional and Theoretic Study of Clitic Cluster Changes in South Slavic”  
(involved the design of a Web crawler to collect linguistically relevant data from  
Web sites in Serbia, Croatia, Slovenia, and Macedonia)

Dissertation Committee: Professor D. Terence Langendoen, Chair  
Professor Thomas Bever  
Professor Joan Bresnan (Stanford University)  
Professor Michael Hammond

**Visiting Scholar** (2001)

Center for the Study of Language and Information  
Stanford University, Palo Alto, CA  
Winter, Spring and Summer Quarters 2001

**M.A. – Linguistics** (1999)

University of Arizona, Tucson, AZ

**B.A. – Linguistics** (1996) – Magna Cum Laude

University of California, Davis, CA

Undergraduate thesis: Testing Grammaticality using a Artificial Neural Network

## **Teaching**

- 2005-Present      Visiting Assistant Professor  
Professional Master's in Computational Linguistics Program  
University of Washington  
Duties:  
    Developed and taught *Shallow Methods in NLP* course  
    Developed and teaching *NLP Systems and Applications* course  
    Taught Graduate Seminar on *Lexical Ambiguity*  
    Promoting program through outreach, contacts with local industry  
    Assisting students in securing summer internships in industry  
    Meeting regularly with students regarding thesis projects
- 2002-Present      Assistant Professor  
Department of Linguistics, California State University, Fresno  
  
2002-2005 *Courses*  
Introduction to Language, Ling 10  
Graduate Natural Language Processing Seminar, Ling 230T  
Undergraduate Computational Linguistics, Ling 152  
Undergraduate Seminar on Cognitive Science, Ling 140T  
  
Average student evaluation for all courses taught: 4.75 (out of 5)
- 1998, 2000      Graduate Teaching Assistant  
Ling 438/538 Computational Linguistics, University of Arizona  
Instructor: D. Terence Langendoen  
Designed syllabus, and assisted in teaching the course.

## **Curriculum Development**

- 2005-Present      Developing new courses for the new Professional Master's in  
Computational Linguistics program at the University of Washington,  
including Linguistics 570 – *Shallow Methods in Statistical Natural  
Language Processing*.
- 2003-2005      Chaired committee to develop two new courses in Cognitive Science  
(CgSci 100 – Foundations of Cognitive Science and CgSci 101 –  
Cognitive Science Seminar), a new minor in Cognitive Science, and a  
new major in Cognitive Science. The minor and the two courses have  
been approved. The major is under review.
- 2002-2003      Developed new curriculum for an undergraduate option in  
Computational Linguistics. Curriculum approved and implemented.
-

### **Industry Experience**

#### **President and Founder (1983-1997)**

SFL Services, Inc.

Sacramento, CA

Focused on systems' integration, with a specialization in the design of interfaces across platforms (mainframes and PC's) and data conversion. Wrote the materials for over 50 classes, which were taught by staff and me. Clients included United Parcel Service, Fidelity National Financial, Pacific Gas and Electric, Enron, and numerous State of California agencies and departments.

### **Publications**

*Refereed Journal  
Publications*

Farrar, Scott and William Lewis. (to appear) "The GOLD Community of Practice: An Infrastructure for Linguistic Data on the Web." *Language Resources and Evaluation*.

Vigliocco, Gabriella, David Vinson, William Lewis, and Merrill Garrett (2004). "Representing the Meaning of Object and Action Words: The Featural and Unitary Semantic Space (FUSS) Hypothesis." *Cognitive Psychology* 48(4) 422-488.

Gerken, LouAnn, Rachel Wilson, and William Lewis (2005). "17-Month-Olds Can Use Distributional Cues to Form Syntactic Categories." *Journal of Child Language* 32, 249-268.

*Publications  
in preparation or in  
submission*

Lewis, William, Scott Farrar, and D. Terence Langendoen (in submission). "Linguistics in the Internet Age: Tools and the Fair Use of Digital Data."

Agbayani, Brian, Chris Golston, and William Lewis (in preparation and testing). "Introduction to Language and Linguistics." Interactive E-text.

*Other Publications*

*(Conference Proceedings, etc.)*

Lewis, William D. (to appear). "ODIN: A Model for Adapting and Enriching Legacy Infrastructure." Proceedings of the e-Humanities Workshop, e-Science 2006, Amsterdam.

Jinguji, Dan, William Lewis, Efthimis N. Efthimiadis, Joshua Minor, Albert Bertram, Shauna Eggers, Joshua Johanson, Brian Nisonger, Ping Yu, and Zhengbo Zhou (to appear). "The University of Washington's UW<sub>CLMAQA</sub> System." In Proceedings of the Text Retrieval Conference (TREC) 2006, Gaithersburg, Maryland.

Lewis, William, Fei Xia, and Dan Jinguji (2006). "Enriching Language Data through Projected Structures." Proceedings of the Texas Linguistics Society 10: Computational Linguistics for Less-Studied Languages.

Lewis, William, Scott Farrar and D. Terence Langendoen (2006). "Linguistics in the Internet Age: Tools and Fair Use." Proceedings of the EMELD'06 Workshop on Digital Language Documentation: Tools and Standards: The State of the Art, Lansing, MI, June 20-22, 2006.

Farrar, Scott and William D. Lewis (2005). "The GOLD Community of Practice: An Infrastructure of Linguistics Data on the Web." Proceedings of the EMELD Workshop on Linguistic Ontologies and Data Categories for Language Resources, Cambridge, Mass., July 2005.

Simons, Gary, William Lewis, Scott Farrar, D. Terence Langendoen, Brian Fitzsimons, and Hector Gonzalez (2004). "The Semantics of Markup: Mapping Legacy Markup Schemes to a Common Semantics." Proceedings of the XMLNLP Workshop, Association for Computational Linguistics, Barcelona, Spain, July 2004.

Simons, Gary, Brian Fitzsimons, D. Terence Langendoen, William Lewis, Scott Farrar, Alexis Lanham, Ruby Basham, and Hector Gonzalez (2004). "A Model for Interoperability: XML Documents as an RDF Database." Proceedings of the EMELD Workshop on Linguistic Databases and Best Practice, Detroit, Michigan, July 2004.

- Lewis, William (2003). "Mining and Migrating Interlinear Glossed Text." Proceedings of the EMELD Workshop on Digitizing & Annotating Texts and Field Recordings, LSA Institute, Michigan State University, July 11th-13th, 2003. <http://emeld.org/workshop/2003/papers03.html>.
- Langendoen, D. Terence, Scott Farrar, William Lewis (2002). "Bridging the Markup Gap: Smart Search Engines for Language Researchers." Proceedings of the Workshop on Resource and Tools for Field Linguistics, May 26-27, Las Palmas, Canary Islands, Spain.
- Farrar, Scott, William Lewis, and D. Terence Langendoen (2002). "A Common Ontology for Linguistic Concepts." Proceedings of the Knowledge Technologies Conference, March 10-13, Seattle.
- Lewis, William. (2002). Measuring Conceptual Distance Using WordNet: The Design of a Metric for Measuring Semantic Similarity. In R. Hayes, W. Lewis, E. Obryan, and T. Zamuner (Eds.), *The University of Arizona Working Papers in Linguistics*. Tucson: University of Arizona.
- Lewis, William (2000). Clitics, Scrambling and Parsing. In J. Weinberg, E. O'Bryan, L. Moll, and J. Haugen (Eds.), *Coyote Papers: The University of Arizona Working Papers in Linguistics*. Tucson: University of Arizona.
- Lewis, William, Scott Farrar, and D. Terence Langendoen (2001). "Building a Knowledge Base of Morphosyntactic Terminology." In S. Bird, P. Buneman, and M. Liberman (Eds.) Proceedings of the IRCS Workshop on Linguistic Databases, 11-13 December 2001, pp. 150-156.
- Lewis, William (2005). *Preserving Endangered Languages' Data: Standards for Data Endurance and Long-term Utility*. Invited talk, October 14<sup>th</sup>, University of Victoria, British Columbia, Canada.

*Invited  
Presentations*

Lewis, William (2005). *Preserving Endangered Languages' Data: Standards for Data Endurance and Long-term Utility*. Invited talk, October 7<sup>th</sup>, University of British Columbia, British Columbia, Canada.

*Presentations*

Lewis, William (2006). *Locating, Recognizing and Enriching Interlinear Glossed Text on the Web*. Paper presented at the University of Washington/Microsoft Computational Linguistics Symposium, held Feb. 3<sup>rd</sup>, Redmond, WA.

Lewis, William (2006). *Locating and Recognizing Interlinear Glossed Text on the Web*. Paper presented at the Linguistics Society of America annual meeting. Albuquerque, NM, January 5th-8th, 2006.

Buszard-Welcher, Laura, Susan Hooyenga, and William Lewis (2005). "The Rosetta Project: All Language Archive and the Impact of the NSDL Through Daughter Repository Networks." National Science Digital Libraries Annual Conference 2005, Denver, Colorado.

Lewis, William (2005). "Taking Resources to the Brink of Interoperation: Profiles, Termsets and Best Practice Markup." Paper presented at the Workshop on Linguistic Ontologies and Data Categories for Language Resources, Cambridge, Mass, July 1st-3rd, 2005. <http://emeld.org/workshop/2005/program.html>.

Lewis, William (2005). "The Online Database of Interlinear Text (ODIN)". Poster presented at the Workshop on Linguistic Ontologies and Data Categories for Language Resources, Cambridge, Mass, July 1st-3rd, 2005. <http://emeld.org/workshop/2005/program.html>.

Jackson, Scott and William D. Lewis (2005). *The Relation Between Prosody and Logical Scope Varies by the Operator*. Poster presented at the Eighteenth Annual CUNY Conference on Human Sentence Processing. Tucson, Arizona, March 31st-April 2nd, 2005.

Lewis, William and Chris Golston (2005). *The Randomness of the Signifier*. Paper presented at the Linguistics Society of

America annual meeting. Oakland, CA, January 6th-9th, 2005.

Eggers, Shauna, D. Terence Langendoen, and William Lewis (2004). *The GOLD Ontology*. Paper presented at the EMELD Workshop on Linguistic Databases and Best Practice, Detroit, Michigan, July 2004.

Lewis, William (2004). *Resource Conversion*. Paper presented at the Linguistics Society of America annual meeting, as part of the Symposium entitled *Endangered Data vs. Enduring Practice: Creating Linguistic Resources That Last*. Boston, MA, January 8th-11th, 2004.

Lewis, William (2001). *Antigrammar: The Role of Negation in Grammar*. In J. Haugen & R. Kennedy (Eds.), *Proceedings of SCIL 2000*. Cambridge, MA: MIT Press.

Lewis, William (2001, October). *Measuring Conceptual Distance: Methods for Measuring the Semantic Similarity of Word Substitution Pairs*. Invited talk, University of Southern Maine, Portland, ME.

Lewis, William, Merrill Garrett and Jason Barker (2001, September). *Measuring Conceptual Distance: The Design of a Metric for Measuring the Semantic Similarity of Word Substitution Pairs*. Poster presented at the 7th Annual Conference on Architectures and Mechanisms for Language Processing (AMLaP): Saarbrücken, Germany.

Lewis, W., and Bever, T. (2000, March). *Structure and Quantifier Interpretation*. Poster presented at the 13th Annual CUNY Human Sentence Processing Conference: La Jolla, CA.

Lewis, William (2000, March). *Antigrammar*. Poster presented at the 13th Annual CUNY Human Sentence Processing Conference: La Jolla, CA.

Gillman, Duffy, William Lewis, and D. Terence Langendoen (2000, March). *The Java Computational Linguistics Environment (JaCLE)*. Poster presented at the 13th Annual CUNY Human Sentence Processing Conference: La Jolla, CA.

### **Professional Activities**

#### *Additional Conferences/Workshops Attended*

Hosted workshop on the General Ontology for Linguistic Description (GOLD) held at the California State University, Fresno, November 12-13, 2004.

Workshop on Designing a Professional Master's Degree Program in Human Language Technology, University of Arizona, Tucson, November 23<sup>rd</sup>, 2002.

EMELD/Linguist List Workshop on ontologies and markup, University of Arizona, Tucson, March 26<sup>th</sup>-28<sup>th</sup>, 2003.

Workshop on Digitizing & Annotating Texts and Field Recordings, LSA Institute, Michigan State University, July 11<sup>th</sup>-13<sup>th</sup>, 2003.

#### *Academic Program Review*

On the review committee for the Academic Program Review of the Graduate Interdisciplinary Program in Cognitive Science at the University of Arizona, Tucson, Arizona. April 2004.

#### *Conference/Workshop Review*

On the Program Review Committee for the Texas Linguistics Society 10 Conference, to be held in Austin, Texas, October 2006.

On the Program Review Committee for the 2006 *CUNY Human Sentence Processing Conference*, held in New York City, New York in March, 2006.

On the Program Review Committee for the 2005 *CUNY Human Sentence Processing Conference*, held in Tucson, Arizona in March, 2005.

On the Program Review Committee for the *Registry of Linguistic Data Categories within an Integrated Language Resource Repository Area Workshop*, held in conjunction with the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, May 2004.

## **Grants**

- 2006* Building a Large Database of Interlinear Text, submitted to NSF Linguistics, under review. (William Lewis, Fei Xia, Bongki Moon PIs, NSF # 0649937, amount \$249,366).
- 2004* Supplement to Data Driven Linguistic Ontology Development Grant, submitted to NSF Linguistics, September 2004, awarded December 2004. (William Lewis PI, NSF BCS-0411348, supplement amount \$22,378.
- Data-Driven Linguistic Ontology Development, Special Grant for Exploratory Research, submitted to NSF Linguistics, December 2003, awarded March 2004. (William Lewis PI, NSF BCS-0411348, \$49,275)
- Cognitive Science Program Development Grant. Internal California State University Fresno grant. Co-written with Dr. Vida Samian, Associate Dean, Arts and Humanities. Submitted Fall 2003, awarded Spring 2004.
- 2003* Mining Interlinear Linguistic Data. PI. Internal California State University Fresno grant. Submitted Spring 2003, awarded Fall 2003.

## **Service**

- Linguistics and Scientific Community* Fall 2003 – Present  
Developed (with student help) and continue to maintain the Online Database of INterlinear text (ODIN), which provides a valuable search facility for language data for the field of linguistics, covering data for 725 of the world's languages.  
URL: <http://www.csufresno.edu/odin>
- Summer-Fall 2006  
On the program committee for the Texas Linguistics Society Conference 10, to be held in Austin, TX, in November 2006
- Spring 2006  
On the review committee for the CUNY Human Sentence Processing Conference held in New York City, NY in March 2006.
- Spring 2005  
On the review committee for the CUNY Human Sentence Processing Conference held in Tucson, Arizona in March, 2005.
- Spring 2004  
Served as an external grant proposal reviewer for the U.S. Civilian Research and Development Foundation.
- Spring 2004  
On the program review committee for the Workshop on A Registry of Linguistic Data Categories within an Integrated Language, held in conjunction with the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, Portugal, May 26<sup>th</sup>-28<sup>th</sup>, 2004.
- April 27th-28th, 2004  
Served on the external review committee for the Academic Program Review (APR) of the University of Arizona Cognitive Science Program.

- Department*                      Fall 2002 – Summer 2005  
Chair, Computational Linguistics Curriculum Committee
- Fall 2002 – Summer 2005  
Member, Linguistics Curriculum Committee
- May 7th, 2005  
Organized and promoted *An Afternoon on Computational Linguistics and Natural Language Processing* at Fresno State, promoting the new major in Computational Linguistics. A guest speaker from Microsoft gave a talk on career opportunities in Computational Linguistics, and various faculty from CSU Fresno gave talks on various Computational Linguistic topics.
- University*                              Fall 2003 – Summer 2005  
Chair, Cognitive Science Committee  
Responsible for overseeing and directing development of the Cognitive Science minor, the Cognitive Science major, and two new Cognitive Science courses. Approval for the minor and the new courses expected in Fall, 2004.
- Spring 2004  
Organized Cognitive Science Lecture Series.
- Fall 2004, Fall 2003  
Elected to Academic Senate (expired Spring 2005)
- Fall 2002  
Formed, with Professors Chris Golston, Brian Agbayani, and Lorin Lachs (Psychology), the Cognitive Science Group. Purpose: To discuss papers and issues relevant to Cognitive Science with students and other faculty.
- Fresno Community*                      Spring 2003 – Spring 2005  
Board Member, Center for Advanced Research and Technology