

Fei Xia

Department of Linguistics
University of Washington
PO Box 352425
Seattle, WA 98195

Tel: (206) 543-9764 (office)

Email: fxia@uw.edu

<http://faculty.washington.edu/fxia>

EMPLOYMENT

- 05/2011-present *Department of Biomedical Informatics and Medical Education
University of Washington, Seattle, Washington*
Adjunct faculty
- 09/2005-present *Department of Linguistics, University of Washington, Seattle, Washington*
9/2015-present: Professor
9/2011-9/2015: Associate Professor
9/2005-9/2011: Assistant Professor
- 10/2001-08/2005 *IBM T. J. Watson Research Center, Yorktown Heights, New York*
10/2002-08/2005: Research Staff Member
10/2001-10/2002: PostDoc
- 05/1996 – 08/1996 *AT&T Bell Laboratories, Murray Hill, New Jersey*
Summer Intern

EDUCATION

- 12/2001 *Ph.D. in Computer Science* University of Pennsylvania
Dissertation: Automatic Grammar Generation from Two Different Perspectives
Advisors: Drs. Martha Palmer and Aravind Joshi
Committee: Steven Bird, Chu-Ren Huang, Tony Kroch, Mitch Marcus, and Vijay Shanker
- 12/1997 *M.S. in Computer Science* University of Pennsylvania
- 07/1992 *B.S. in Computer Science* Peking University

GRANTS

1. NSF grant (BCS-1561833): co-PI (PI: Emily Bender)
 - Title: “Automatic grammar engineering for endangered languages based on cross-linguistic resources”
 - Funding period: 08/15/2016 – 01/31/2020
 - UW amount: \$429,864
 - Goal: The goal of the project is to design a new approach for automatically generating computational grammars by combining computational techniques developed for well-studied languages, data collected and annotated by field linguists, and a cross-linguistic grammar resource (the LinGO Grammar Matrix).

2. NIH grant (1R21EB016872), co-PI (PI : Meliha Yildiz)
 - Title: *Using NLP to Extract Clinically Important Recommendations from Radiology Reports*
 - Funding period: 3/1/2014-2/28/2016
 - UW amount: \$474,800
 - Goal : The goal of the project is to design and implement a natural language processing system that identifies the clinically important recommendation information in free-text radiology notes. The extracted information will be integrated to EMR to improve visibility as well as to monitor recommendation follow-ups.

3. NSF grant (BCS-1160274): co-PI (PI: Emily Bender)
 - Title: “AGGREGATION: Automatic Generation of Grammars for Endangered Languages from Glosses and Typological Information”
 - Funding period: 09/15/2012 - 02/28/2015
 - UW amount: \$224,039
 - Goal: The world's linguistic diversity is diminishing at an alarming rate, and there are not enough resources (trained field linguists or funding for them) to document all the endangered languages before they are gone. Thus there is a critical need for software tools to support the efficiency of field linguists. This project will develop software tools to assist in the documentation of endangered languages by merging two types of resources: Collections of linguistic examples curated by linguists and a cross-linguistic computational grammar resource, called the Grammar Matrix. The result will be a system for creating machine-readable, or implemented, grammars from data collected and annotated by field linguists.

4. UW Royalty Research Fund (RRF) grant: co-investigator (PI: Meliha Yetisgen-Yildiz)
 - Title: “Annotating and Detecting Medical Events in Clinical Data”
 - Funding period: 09/01/2012-08/30/2014
 - UW Amount: \$39,968
 - Goal: Over the last decade, Electronic Medical Record (EMR) systems have become increasingly integral to the provision of health care services. Accessibility to the details of patient data in EMR systems is critical to improve the health care process and advance clinical research. The primary goal of the proposed project is to study medical events and relations between the events, and then annotate such events and relations in a clinical corpus. The corpus will be used to initiate the NLP research in our institution on developing text-processing approaches for medical event extraction.

5. IARPA grant (D11PC20153): subcontractor (PI: Kathleen McKeown at Columbia Univ)
 - Title: “Discovering and Explaining Technical Emergence through Analysis of the Language and Structure of Scientific Publications”
 - Funding period: 8/08/2011-2/07/2013
 - UW Amount: \$297,872
 - Goal: The goal of the project is to design an integrated system that automatically detects technical emergence in the scientific literature. It identifies indicators from full text scientific documents, predicts technical emergence, and justifies nominated technical topics. This is a collaborative project with Columbia University, University of Maryland, University of Michigan, and Cambridge University.

6. Microsoft Research Grant: collaborator (PI: Meliha Yetisgen-Yildiz)
 - Title: Extracting Critical Illness Phenotypes from Electronic Medical Records
 - Funding period: 1/1/2011-5/31/2013
 - UW Amount: \$250,000
 - Goal: The main research aim of this project is to design and develop a text processing platform to identify critical illness phenotypes with associated indicators in EMR. The platform will be released to research community as an open-source tool. This is a joint project with Microsoft Research Natural Language Processing group.

7. NSF Workshop Grant (IIS-1027289): PI
 - Title: Workshop on NLP and Linguistics: Finding the Common Ground
 - Funding period: 10/01/2010-09/30/2011
 - UW Amount: \$16,996
 - Goal: The grant is used to support the workshop on “NLP and Linguistics: Finding the Common Ground” in conjunction with ACL 2010 at Uppsala, Sweden. The goal of the workshop is to carefully examine the relationship between linguistics and NLP and determine how incorporating linguistic knowledge into NLP systems can advance the state of the art of NLP and how NLP can assist linguistic studies through automatic collection and analysis of linguistic data .

8. NIH/NLM K99/R00 Pathway to Independence Award (1K99LM010227-0110): Consultant (PI: Imre Solti)
 - Title: Increasing Clinical Trial Enrollment: A Semi-Automated Patient Centered Approach
 - Funding period: 09/30/2009-09/29/2011 (Phase 1) and 09/30/2011-09/29/2014 (Phase 2)
 - UW Amount: \$168,340 (Phase I) and \$741,922 (Phase II)
 - Goal: (a) To identify concepts that overlap between the electronic medical record's (EMR) clinical notes and the free text of clinical trial announcements, (b) To create an algorithm that automatically generates questions to request information directly from the patient if the information is not available or accessible in the records, (c) To evaluate the in-vitro, laboratory performance of the application

9. NSF REU Supplement (IIS-0939733): PI
 - Title: Supplement to the NSF CAREER Grant
 - Funding period: 7/14/2009 -- 5/30/2013

- UW amount: \$15,925
- Goal: The grant is used to support undergraduate student research in the RiPLEs project.

10. NSF CAREER grant (BCS-0748919): PI

- Title: Information Engineering and Synthesis for Resource-poor Languages
- Funding period: 6/15/2008 -- 5/30/2013
- UW Amount: \$500,000 (\$614,532)
- Goal: This project has two intertwined goals. The first goal is to create a framework that allows the rapid development of resources for resource-poor languages (RPLs). We will accomplish this goal by bootstrapping NLP tools with initial seeds created by projecting syntactic information from resource-rich languages to RPLs. The second goal is to use the automatically created resources to perform cross-lingual study on a large number of languages to discover linguistic knowledge. The knowledge will not only deepen our understanding on languages, but also provide additional information that can be incorporated into the bootstrapping module to produce better NLP tools.

11. NSF Hindi grant (CNS-0751213): PI

- Title: Collaborative research: A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu
- Funding period: 5/1/2008 -- 4/30/2014
- UW Amount: \$196,001
- Goal: We are collaborating with the Univ. of Colorado, Columbia Univ, the Univ. of Massachusetts at Amherst and IIIT-Hyderabad to build a Hindi/Urdu treebank that will be a second-generation treebank: a multi-representational and multi-layered Treebank. The Treebank consists of phrase structure, dependency structure, and predicate-argument structure. The emphasis on a multi-representational and multi-layered treebank will help move research on parsing to a deeper understanding of the interaction of computation, learning, and syntax.

12. NSF GOLDcomm grant (BCS-0720670), co-PI (PI : Scott Farrar)

- Title : Implementing the GOLD Community of Practice: Laying the Foundations for a Linguistics Cyberinfrastructure
- Funding period : 9/1/2007-8/30/2010
- UW Amount: \$147,456
- Goal : The primary goal of this project is the implementation of the GOLD Community of Practice, a Web architecture for linking on-line linguistic data to linguistic knowledge captured by the General Ontology for Linguistic Description (GOLD). The ever increasing mass of linguistic data available on the Web and the advances in best practice data standards will be brought together to realize this model. A successful implementation could pave the way for a future cyber-infrastructure for linguistics.

13. NSF planning grant (CNS-0708719), PI.

- Title: General Techniques for Creating Treebanks with Multiple Representation
- Funding period: 5/01/2007-4/30/2008
- UW Amount: \$20,813
- Goal: The goal of this project is to study the feasibility of creating a multi-representational treebank with both dependency structure and phrase structure and of automatically converting

dependency structure into phrase structure. The success of the project has led to the embark of the Hindi/Urdu Treebank Project.

14. UW Royalty Research Fund (RRF) grant (#65-1337), PI.

- Title : Towards automatic enrichment and analysis of linguistic data for threatened and endangered languages
- Funding period : 3/16/2007-3/15/2009
- UW Amount : \$39,678
- Goal : The central theme of this project is to explore the automation of the language analysis and resource creation process. To be more specific, we propose building a system that automatically projects syntactic structures from annotated data for resource-rich languages (such as English) to data for digitally impoverished languages (such as Hausa), where the resulting projected structures can be used for the development of language-specific tools (e.g., part-of-speech taggers, parsers, machine translation systems). The purpose of this proposal is to prove the technology on ten languages, and the subsequently apply for external funding to extend the methods to a much larger set of languages.

15. UW Student Technology Fee (STF) grant (2006-058-1), Faculty lead.

- Title : Natural Language Computing Cluster
- Funded in 2006
- UW Amount : \$66,710
- Goal : The fund was used to purchase computer nodes for the cluster in the Computational Linguistics Lab at the UW Linguistics Department

PUBLICATIONS

(1) Peer-reviewed Journal Articles: Full Paper Review

- [1] Fei Xia, William D. Lewis, Michael W. Goodman, Glenn Slayden, Ryan Georgi, Joshua Crowgey, and Emily Bender, 2016. “Enriching a Massively Multilingual Database of Interlinear Glossed Text”, *Journal of Language Resources and Evaluation (LRE)*, 50(2): 321-349.
- [2] Michael Wayne Goodman, Joshua Crowgey, Fei Xia, and Emily M. Bender, 2015. “Xigt: Extensible Interlinear Glossed Text for Natural Language Processing”, *Journal of Language Resources and Evaluation (LRE)*, 49(2), pp 455-485.
- [3] Ryan Georgi, Fei Xia, and William D. Lewis, 2014. “Capturing Divergence in Dependency Trees to Improve Syntactic Projection”, *Journal of Language Resources and Evaluation (LRE)*, 48(4), pp 709-739.
- [4] Meliha Yetisgen-Yildiz, Martin Gunn, Fei Xia, and Tom Payne, 2013. “Text Processing Pipeline to Extract Recommendations from Radiology Reports”, *Journal of Biomedical Informatics (JBI)*, 46(2):354-362.

- [5] Cosmin Adrian Bejan, Lucy Vanderwende, Fei Xia, Meliha Yetisgen-Yildiz, 2013. "Assertion modeling and its role in clinical phenotype identification", *Journal of Biomedical Informatics (JBI)*, 2013; 46(1):68-74.
- [6] Louise Deleger, Katalin Molnar, Guergana Savova, Fei Xia, Todd Lingren, Qi Li, Keith Marsolo, Anil G. Jegga, Megan Kaiser, Laura Stoutenborough, and Imre Solti, 2013. Large Scale Evaluation of Automated Clinical Note De-identification and its Impact on Information Extraction. *Journal of the American Medical Informatics Association (JAMIA)*, 20(1): 84-94.
- [7] Cosmin Adrian Bejan, Fei Xia, Lucy Vanderwende, Mark M. Wurfel, and Meliha Yetisgen-Yildiz, 2012. "Pneumonia identification using statistical feature selection", *Journal of American Medical Informatics Association (JAMIA)*, 19(5): 817-823.
- [8] Scott Russell Halgrim, Fei Xia, Imre Solti, Eithon Cadag, Ozlem Uzuner, 2011. "A cascade of MaxEnt classifiers applied to extracting medication information from discharge summaries", *Journal of Biomedical Semantics* 2011, 2 (Suppl 3):S2.
- [9] Ozlem Uzuner, Imre Solti, Fei Xia, and Eithon Cadag, 2010. "Community Annotation Experiment for Ground Truth Generation for the i2b2 Medication Challenge", *Journal of the American Medical Informatics Association (JAMIA)*, 17:519-523.
- [10] William Lewis and Fei Xia, 2010. "Developing ODIN: A Multilingual Repository of Annotated Language Data for Hundreds of the World's Languages," *Journal of Literary and Linguistic Computing (LLC)*, 25(3):303-319.
- [11] Michael Tepper and Fei Xia, 2010. "Inducing Morphemes Using Light Knowledge," *Journal of ACM Transactions on Asian Language Information Processing (TALIP)*, 9(1): 1-38.
- [12] Fei Xia, Martha Palmer and K. Vijay-Shanker, 2005. "Automatically Generating Tree Adjoining Grammars from Abstract Specifications," *Journal of Computational Intelligence*, 21(3), 246-287.
- [13] Nianwen Xue, Fei Xia, Fu-dong Chiou, and Martha Palmer, 2005. "The Penn Chinese Treebank: Phrase Structure Annotation of a Large Corpus," *Journal of Natural Language Engineering*, 11(2): 207-238.

(2) Peer-reviewed Book Chapters: Full Paper Review

- [14] Riyaz Ahmad Bhat, Rajesh Bhatt, Annahita Farudi, Prescott Klassen, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, Ashwini Vaidya, Sri Ramagurumurthy Vishnu, and Fei Xia, 2014. "The Hindi/Urdu Treebank Project", to appear in the *Handbook of Linguistics Annotation* (edited by Nancy Ide and James Pustejovsky), Springer Press.
- [15] Fei Xia and Martha Palmer, 2010. "From Treebank to Tree-Adjoining Grammar", in "*Supertagging: Using Complex Lexical Descriptions in Natural Language Processing*", edited by Srinivas Bangalore and Aravind K. Joshi, pages 35-72, MIT Press.
- [16] Fei Xia, Martha Palmer, and Vijay Shanker, 2010. "Developing Tree-Adjoining Grammars with Lexical Descriptions," in "*Supertagging: Using Complex Lexical Descriptions in Natural Language Processing*", edited by Srinivas Bangalore and Aravind K. Joshi, pages 73-110, MIT Press.

- [17] Martha Palmer, Chung-hye Han, Fei Xia, Dania Egedi and Joseph Rosenzweig, 2000. “Constraining Lexical Selection across Languages Using Tree Adjoining Grammars,” in *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, a CSLI volume edited by Anne Abeille and Owen Rambow, pages 445-466.
- [18] Christy Doran, Beth Ann Hockey, Anoop Sarkar, Srinivas Bangalore and Fei Xia, 2000. “Evolution of the XTAG System,” in *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, a CSLI volume edited by Anne Abeille and Owen Rambow, pages 371-404.

(3) Peer-reviewed Conference Proceedings: Full Paper Review

- [19] Ryan Georgi, Michael Wayne Goodman, and Fei Xia, 2016. “A Web-framework for ODIN Annotation”, in Proceedings of ACL-2016 System Demonstrations, pp 31-36, Aug 7-10, Berlin, Germany.
- [20] Prescott Klassen, Fei Xia, and Meliha Yetisgen, 2016. “Annotating and Detecting Medical Events in Clinical Notes”, in Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016), May 23-28, Portoroz, Slovenia.
- [21] Maria Antoniak, Eric Bell, and Fei Xia, 2015. “Leveraging Paraphrase Labels to Extract Synonyms from Twitter”, in Proceedings of the 28th International Florida Artificial Intelligence Research Society (FLAIRS) Conference, May 18-20, Hollywood, Florida, USA.
- [22] Fei Xia, Michael Wayne Goodman, Ryan Georgi, Glenn Slayden, and William D. Lewis, 2015. “Enriching, Editing, and Representing Interlinear Glossed Text”, in Proceedings of the 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2015), April 14-20, Cairo, Egypt.
- [23] Xuezhe Ma and Fei Xia, 2014. “Unsupervised Dependency Parsing with Transferring Distribution via Parallel Guidance and Entropy Regularization”, in Proceedings of ACL-2014, Baltimore, MD.
- [24] Yan Song and Fei Xia, 2014. “Modern Chinese Helps Archaic Chinese Processing: Finding and Exploiting the Shared Properties”, in Proceedings of LREC 2014, Reykjavik, Iceland.
- [25] Fei Xia, William Lewis, Michael Wayne Goodman, Joshua Crowgey and Emily M. Bender, 2014. “Enriching ODIN”, in Proceedings of LREC 2014, Reykjavik, Iceland.
- [26] Prescott Klassen, Fei Xia, Lucy Vanderwende and Meliha Yetisgen, 2014. “Annotating Clinical Events in Text Snippets for Phenotype Detection”, in Proceedings of LREC 2014, Reykjavik, Iceland.
- [27] Kam Tang Lau, Yan Song, and Fei Xia, 2013. “The Construction of a Segmented and Part-of-speech Tagged Archaic Chinese Corpus: A Case Study on Huainanzi”, in Proceedings of the 12th China National Conference on Computational Linguistics (CNCCL 2013), Oct 10-12, Suzhou, China.
- [28] Yan Song and Fei Xia, 2013. “A Common Case of Jekyll and Hyde: the Synergistic Effect of Using Divided Source Training Data for Feature Augmentation”, in Proceedings of IJCNLP, Oct 14-18, Nagoya, Japan.

- [29] Xuezhe Ma and Fei Xia, 2013. "Dependency Parser Adaptation with Subtrees from Auto-Parsed Target Domain Data", short paper, In Proceedings of ACL, Sofia, Bulgaria, Aug 2013.
- [30] Ryan Georgi, Fei Xia, and William D. Lewis, 2013. "Enhanced and Portable Dependency Projection Algorithms Using Interlinear Glossed Text", short paper, In Proceedings of ACL, Sofia, Bulgaria, Aug 2013.
- [31] Ryan Georgi, Fei Xia, and William D. Lewis, 2012. "Improving Dependency Parsing with Interlinear Glossed Text and Syntactic Projection", short paper, In Proceedings of COLING. Mumbai, India, Dec 2012.
- [32] Yan Song, Prescott Klassen, and Fei Xia, 2012. "Entropy-based Training Data Selection for Domain Adaptation", short paper, In Proceedings of COLING. Mumbai, India, Dec 2012.
- [33] Dong Wang and Fei Xia, 2012. "Effort of Genre Variation and Prediction of System Performance", In Proceedings of LREC, Istanbul, Turkey, May 22-25, 2012.
- [34] Yan Song and Fei Xia, 2012. "Using a Goodness Measurement for Domain Adaptation: A Case Study on Chinese Word Segmentation", In Proceedings of LREC, Istanbul, Turkey, May 22-25, 2012.
- [35] Ryan Georgi, Fei Xia, and William D. Lewis. 2012. "Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms", In Proceedings of LREC, Istanbul, Turkey, May 22-25, 2012.
- [36] Michael Tepper, Daniel Capurro, Fei Xia, Lucy Vanderwende, and Meliha Yetisgen-Yildiz, 2012. Statistical Section Segmentation in Free-Text Clinical Records. In the Proceedings of the LREC, Istanbul, Turkey, May 22-25, 2012.
- [37] Rajesh Bhatt, Owen Rambow, and Fei Xia, 2011. "Linguistic Phenomena, Analyses, and Representations: Understanding Conversion between Treebanks", In the Proc. of the IJCNLP, Chiang Mai, Thailand, Nov 9-13, 2011.
- [38] Cuijun Wu, Fei Xia, Louise Deleqer, and Imre Solti, 2011. "Statistical Machine Translation for Biomedical Text: Are We There Yet?" In the Proceedings of the AMIA 2011 Annual Symposium, pages 1290-1299, Washington DC, Oct 22-26, 2011.
- [39] Meliha Yetisgen-Yildiz, Martin Gunn, Fei Xia, and Tom Payne, 2011. "Automatic Identification of Critical Follow-Up Recommendation Sentences in Radiology Reports". In the Proc. of the AMIA 2011 Annual Symposium, Washington DC, Oct 22-26, 2011.
- [40] Ryan Georgi, Fei Xia and William Lewis, 2010. "Comparing Language Similarity across Genetic and Typologically-Based Groupings," In Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), pages 385-393, Beijing, China, August 23-27, 2010.
- [41] Alex Cheng, Fei Xia, and Jianfeng Gao, 2010. "A comparison of unsupervised methods for Part of Speech Tagging in Chinese," In Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010), Poster Volume, pages 135-143, Beijing, China, August 23-27, 2010.

- [42] Fei Xia, Carrie Lewis, and Will Lewis, 2010. "The Problems of Language Identification within Hugely Multilingual Data Sets," The Seventh International Conference on Language Resources and Evaluation (LREC-2010), Malta, May 17-23, 2010.
- [43] Archana Bhatia, Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, Michael Tepper, Ashwini Vaidya, and Fei Xia, 2010. "Empty Categories in a Hindi Treebank," The Seventh International Conference on Language Resources and Evaluation (LREC-2010), Malta, May 17-23, 2010.
- [44] Martha Palmer, Rajesh Bhatt, Bhuvana Narasimhan, Owen Rambow, Dipti Misra Sharma, and Fei Xia, 2009. "Hindi Syntax: Annotating Dependency, Lexical Predicate-Argument Structure, and Phrase Structure", The 7th International Conference on Natural Language Processing (ICON-2009), Hyderabad, India, Dec 14-17, 2009.
- [45] Fei Xia, William Lewis and Hoifung Poon, 2009. "Language ID in the Context of Harvesting Language Data off the Web," The 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2009), Athens, Greece, March 30 – April 3, 2009.
- [46] William Lewis and Fei Xia, 2009. "Parsing, Projecting & Prototypes: Repurposing Linguistic Data on the Web," The 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2009), Demo session, Athens, Greece, March 30 – April 3, 2009.
- [47] Fei Xia and William Lewis, 2008. "Repurposing Theoretical Linguistic Data for Tool Development and Search," *The Third International Joint Conference on Natural Language Processing (IJCNLP-2008)*, Hyderabad, India, Jan 7-12, 2008.
- [48] William Lewis and Fei Xia, 2008. "Automatically Identifying Computationally Relevant Typological Features," *The Third International Joint Conference on Natural Language Processing (IJCNLP-2008)*, Hyderabad, India, Jan 7-12, 2008.
- [49] Michael Tepper and Fei Xia, 2008. "A Hybrid Approach to the Induction of Underlying Morphology," *The Third International Joint Conference on Natural Language Processing (IJCNLP-2008)*, Hyderabad, India, Jan 7-12, 2008.
- [50] Fei Xia and William Lewis, 2007. "Multilingual Structural Projection across Interlinearized Text," The Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2007), Rochester, NY, April 22-27, 2007.
- [51] Fei Xia and Michael McCord, 2004. "Improving a Statistical MT System with Automatically Learned Rewrite Patterns," *The 20th International Conference on Computational Linguistics (COLING-2004)*, Geneva, Switzerland, August 22– 29, 2004.
- [52] Christoph Tillmann and Fei Xia, 2003. "A Phrase-Based Unigram Model for Statistical Machine Translation," *The 3rd Human Language Technology Conference (HLT/NAACL-2003)*, Edmonton, Canada, May 27 – June 2, 2003.
- [53] Y. Al-Onaizan, R. Florian, M. Franz, H. Hassan, Y. S. Lee, S. McCarley, K. Papineni, S. Roukos, J. Sorensen, C. Tillmann, T. Ward, F. Xia, 2003. "TIPS: A Translingual Information Processing System", The 3rd Human Language Technology Conference (HLT/NAACL-2003), Demo Session, Edmonton, Canada, May 27 – June 2, 2003.

- [54] Fei Xia, Chung-hye Han, Martha Palmer and Aravind Joshi, 2001. “Automatically Extracting and Comparing Lexicalized Grammars for Different Languages,” *The 17th International Joint conference on Artificial Intelligence (IJCAI-2001)*, Seattle, Washington, August 4 – 10, 2001.
- [55] Fei Xia and Martha Palmer, 2001. “Converting Dependency Structures to Phrase Structures,” *The 1st Human Language Technology Conference (HLT-2001)*, San Diego, CA, March 18 – 21, 2001.
- [56] Fei Xia, Martha Palmer, and Aravind Joshi, 2000. “A Uniform Method of Grammar Extraction and Its Applications,” *The Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*, Hong Kong, October 7 – 8, 2000.
- [57] Fei Xia, Martha Palmer, Nianwen Xue, Mary Ellen Okurowski, John Kovarik, Fu-Dong Chiou, Shizhe Huang, Tony Kroch, and Mitch Marcus, 2000. “Developing Guidelines and Ensuring Consistency for Chinese Text Annotation,” *The 2nd International Conference on Language Resources and Evaluation (LREC-2000)*, Athens, Greece, May 31 – June 2, 2000.
- [58] Fei Xia, 1999. “Extracting Tree Adjoining Grammars from Bracketed Corpora,” *The 5th Natural Language Processing Pacific Rim Symposium (NLPRS-1999)*, Beijing, China, November 5 – 7, 1999.
- [59] Fei Xia, Martha Palmer, and K. Vijay-Shanker, 1999. “Towards Semi-automating Grammar Development,” *The 5th Natural Language Processing Pacific Rim Symposium (NLPRS-1999)*, Beijing, China, November 5 – 7, 1999.
- [60] Hiyun Alshawi, Adam Buchsbaum, and Fei Xia, 1997. “A Comparison of Head Transducers and Transfer for a Limited Domain Translation,” in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-1997)*, Madrid, Spain, July 7 – 11, 1997.
- [61] Chung-hye Han, Fei Xia, Martha Palmer, and Joseph Rosenzweig, 1996. “Capturing Language Specific Constraints on Lexical Selection with Feature-Based LTAGs,” *The International Conference on Chinese Computing (ICCC-1996)*, Singapore, June 4 – 7, 1996.

(4) Peer-reviewed Workshop Proceedings: Full Paper Review

- [62] Ryan Georgi, Fei Xia, and William D. Lewis, 2015. “Enriching Interlinear Text using Automatically Constructed Annotators”, in *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH-2015)*, in conjunction with ACL 2015, July 30, Beijing, China.
- [63] Emily M. Bender, Joshua Crowgey, Michael Wayne Goodman, and Fei Xia, 2014. “Learning Grammar Specifications from IGT: A Case Study of Chintang”, in *Proceedings of the Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL)*, in conjunction with ACL 2014, June 26, Baltimore, Maryland, USA.
- [64] Emily M. Bender, Michael Wayne Goodman, Joshua Crowgey, and Fei Xia, 2013. “Towards Creating Precision Grammars from Interlinear Glossed Text: Inferring Large-scale Typological Properties”, in *Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2013)*, in conjunction of ACL 2013, Sofia, Bulgaria.
- [65] Michael Tepper, Heather L. Evans, Fei Xia, Meliha Yetisgen-Yildiz. 2013. “Modeling Annotator Rationales with Application to Pneumonia Classification”, in *Proceedings of the 2013 AAAI workshop*

on Expanding the Boundaries of Health Informatics Using Artificial Intelligence (HIAI 2013), July 15, Bellevue, WA.

- [66] Lucy Vanderwende, Fei Xia, and Meliha Yetisgen-Yildiz, 2013. “Annotating Change of State for Clinical Events”, in Proceedings of the 1st Workshop on Events: Definition, Detection, Coreference, and Representation, in conjunction with NAACL-2013, Atlanta, GA.
- [67] Rajesh Bhatt, Owen Rambow, and Fei Xia, 2012. “Creating a Tree Adjoining Grammar from a Multilayer Treebank”, in Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11), pages 162-170, Paris, France, September 2012.
- [68] Fei Xia and Meliha Yetisgen-Yildiz, 2012. “Clinical corpus annotation: challenges and strategies”, in Proceedings of the third Workshop on Building and Evaluating Resources for Biomedical Text Mining, in conjunction with LREC-2012, Istanbul, Turkey.
- [69] Rajesh Bhatt and Fei Xia, 2012. “Challenges in Converting between Treebanks: a Case Study from the HUTB”, in Proceedings of META-RESEARCH Workshop on Advanced Treebanking, in conjunction with LREC-2012, Istanbul, Turkey.
- [70] Meliha Yetisgen-Yildiz, Bradford Glavan, Fei Xia, Lucy Vanderwende, and Mark Wurfel, 2011. “Identifying Patients with Pneumonia from Free-Text Intensive Care Unit Reports”. In Proc. of the ICML workshop on Learning from Unstructured Clinical Text, Bellevue, WA, July 2, 2011.
- [71] Kelly Peterson, Matt Hohensee, and Fei Xia, 2011. “Email Formality in the Workplace: A Case Study on the Enron Corpus,” In Proceedings of the 2011 ACL Workshop on Language in Social Media (LSM 2011), Portland, Oregon, June 23, 2011.
- [72] Scott Halgrim, Fei Xia, Imre Solti, Eithon Cadag, and Ozlem Uzuner, 2010. “Extracting medication information from discharge summaries,” In Proceedings of the Second Louhi Workshop on Text and Data Mining of Health Documents, in conjunction with NAACL 2010, Los Angeles, CA, June 5, 2010.
- [73] Meliha Yetisgen-Yildiz, Imre Solti, Fei Xia, and Scott Halgrim, 2010. “Preliminary Experience with Amazon’s Mechanical Turk for Annotating Medical Named Entities,” In Proceedings of the NAACL workshop on Creating Speech and Text Language Data with Amazon’s Mechanical Turk, Los Angeles, CA, June 6, 2010.
- [74] Imre Solti, Colin Cooke, Fei Xia, and Mark Wurfel, 2009: “Automated classification of radiology reports for acute lung injury: Comparison of keyword and machine learning based natural language processing approaches,” *NLP Workshop, IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2009)*, Washington DC, November 1-4, 2009.
- [75] Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma, Fei Xia, 2009. “A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu,” The Third Linguistic Annotation Workshop (The LAW III) in conjunction with ACL/IJCNLP 2009. Singapore. Aug 6-7, 2009.
- [76] Fei Xia and William Lewis, 2009. “Applying NLP Technologies to the Collection and Enrichment of Language Data on the Web to Aid Linguistic Research,” The workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH-SHELT&R 2009). In conjunction with EACL-2009, Athens, Greece, March 30, 2009.

- [77] Fei Xia, Owen Rambow, Rajesh Bhatt, Martha Palmer, and Dipti Misra Sharma, 2009. "Towards a Multi-Representational Treebank," in Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories (TLT 2009), Groningen, Netherlands, Jan 23-24, 2009.
- [78] Fei Xia, 2008. "The evolution of a statistical NLP course," *In Proceedings of the Third ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, Columbus, Ohio, June 19-20, 2008.
- [79] Emily Bender, Fei Xia, and Erik Bansleben, 2008. "Building a flexible, collaborative, intensive master's program in computational linguistics," *In Proceedings of the Third ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, Columbus, Ohio, June 19-20, 2008.
- [80] Achim Ruopp and Fei Xia, 2008. "Finding parallel texts on the web using cross-language information retrieval", *The Workshop on Cross Language Information Access* in conjunction with IJCNLP-2008. Hyderabad, India, Jan 7-12, 2008.
- [81] William Lewis, Fei Xia, and Dan Jinguji, 2006. "Enriching Language Data through Projected Structures", *The Workshop on Computational Linguistics for Less-studied Languages*, organized by Texas Linguistics Society (TLSX), Austin, Texas, Nov 3-5, 2006.
- [82] Fei Xia and Lap Cheung, 2006, "Features, Bagging, and System Combination for the Chinese POS Tagging Task," *The 5th SIGHAN Workshop on Chinese Language Processing (SIGHAN 2006)*, Sydney, Australia, July 22-23, 2006.
- [83] Fei Xia, Chung-hye Han, Martha Palmer, and Aravind Joshi, 2000. "Comparing Lexicalized Treebank Grammars Extracted from Chinese, Korean, and English Corpora," *The 2nd Chinese Language Processing Workshop (CLP-2000)*, Hong Kong, October 8, 2000.
- [84] Anoop Sarkar, Fei Xia, and Aravind Joshi, 2000. "Some Experiments on Indicators of Parsing Complexity for Lexicalized Grammars," *Efficiency in Large-Scale Parsing Systems Workshop*, Luxembourg, Germany, August 5, 2000.
- [85] Fei Xia and Martha Palmer, 2000. "Evaluating the Coverage of LTAGs on Annotated Corpora," *The Workshop on Using Evaluation within HLT Programs: Results and Trends*, Athens, Greece, May 30, 2000.
- [86] Fei Xia and Tonia Bleam, 2000. "A Corpus-based Evaluation of Syntactic Locality in TAGs," *The 5th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+ 2000)*, Paris, France, May 25 – 27, 2000.
- [87] Fei Xia and Martha Palmer, 2000. "Comparing and Integrating Tree Adjoining Grammars," *The 5th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+ 2000)*, Paris, France, May 25 – 27, 2000.
- [88] Fei Xia, Martha Palmer, K. Vijay-Shanker, and Joseph Rosenzweig, 1998. "Consistent Grammar Development Using Partial-Tree Descriptions for LTAGs," *The 4th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+ 1998)*, Philadelphia, August 1 – 3, 1998.
- [89] Christine Doran, Beth Hockey, Philip Hopely, Joseph Rosenzweig, Anoop Sarkar, Srinivas Bangalore, Fei Xia, Alexis Nasr and Owen Rambow, 1997. "Maintaining the Forest and Burning out the

Underbrush in XTAG,” *the Workshop on Computational Environments for Grammar Development and Language Engineering (ENVGRAM-1997)*, Madrid, Spain, July 12, 1997.

- [90] Hiyan Alshawi and Fei Xia, 1997. “English-to-Mandarin Speech Translation with Head Transducers,” *the Workshop of Spoken Language Translation (SLT-1997)*, Madrid, Spain, July 11, 1997.

(5) Peer-reviewed Abstracts:

- [91] Maria Antoniak, Eric Bell, and Fei Xia, 2015. “Extracting Topic-Specific Synonyms from Twitter”, in Proceedings of the 10th Annual Women in Machine Learning Workshop, in conjunction with NIPS, Montreal, Canada, Dec 7.

- [92] Meliha Yetisgen-Yildiz, Prescott Klassen, Lucy Vanderwende, and Fei Xia, 2014. “A New Corpus for Clinical Events with Change of State”, in Proceedings of the 2014 AMIA Joint Summit on Translational Science, San Francisco, April 7-11.

- [93] Meliha Yetisgen-Yildiz, Cosmin A. Bejan, Prescott Klassen, Michael Tepper, Lucy Vanderwende, and Fei Xia, 2013. “Text Processing Tools from the University of Washington Biomedical Language Processing Group”, in Proceedings of the 2013 AMIA Symposium.

- [94] Emily M. Bender, Fei Xia, Joshua Crowgey, and Michael Wayne Goodman, 2013. “Towards Automatic Detection of Morphosyntactic Systems from IGT”, in Proceedings of the Workshop on Exploring Data from Language Documentation.

- [95] Meliha Yetisgen-Yildiz, Cosmin A. Bejan, Lucy Vanderwende, Fei Xia, Heather L. Evans, and Mark M. Wurfel. 2013. “Automated Tools for Phenotype Extraction from Medical Records”, Abstract in the 2013 AMIA Joint Summits on Translational Science.

- [96] Michael Tepper, Fei Xia, and Meliha Yetisgen-Yildiz, 2012. “Smoking Status Detection across Domains”, in Proceedings of the AMIA Fall Symposium, Chicago, Illinois, November 2012.

- [97] Meliha Yetisgen-Yildiz, Bradford Glavan, Fei Xia, Lucy Vanderwende, and Mark Wurfel, 2011. “Extraction of Pneumonia Cases from Free-Text Intensive Care Unit Reports”. The AMIA 2011 Annual Symposium.

- [98] Imre Solti, Scott Halgrim, and Fei Xia, 2010. “Addressing the Annotation Bottleneck for Clinical Natural Language Processing: Testing the Feasibility of Domain Adaptation for Medical Text,” The AMIA 2010 Annual Symposium, Washington DC, Nov 13-17, 2010.

- [99] Meliha Yetisgen-Yildiz, Imre Solti, and Fei Xia, 2010. “Using Amazon’s Mechanical Turk for Annotating Medical Named Entities,” The AMIA 2010 Annual Symposium, Washington DC, Nov 13-17, 2010.

- [100] Imre Solti, Colin R. Cooke, Fei Xia, and Mark M. Wurfel, 2010. “Peeling Away the Black Box Label: Clinical Validation of a MaxEnt Machine Learning Character N-gram Feature Set for Acute Lung Injury,” 2010 AMIA Summit on Translational Bioinformatics, San Francisco, CA, March 10-12, 2010.

- [101] Scott Russell Halgrim, Fei Xia, Imre Solti, Eithon Cadag, and Ozlem Uzuner, 2010. "Statistical Extraction of Medication Information from Clinical Records," 2010 AMIA Summit on Translational Bioinformatics, San Francisco, CA, March 10-12, 2010.
- [102] Fei Xia, Carrie Lewis, and William Lewis, 2010. "*Language ID for a Thousand Languages*," eLanguage. LSA Annual Meeting Extended Abstracts, Baltimore, Maryland, Jan 7-10, 2010.
- [103] William Lewis, Fei Xia, and Daniel Jinguji, 2007. "Projecting structure onto data for resource-poor and endangered languages," LSA Annual Meeting, Anaheim, CA, 4-7 January 2007.

(6) Technical Reports

- [104] Fei Xia, 2000. "The Segmentation Guidelines for the Penn Chinese Treebank (3.0)," *IRCS Report 00-06*, University of Pennsylvania, October 2000.
- [105] Fei Xia, 2000. "The Part-of-Speech Guidelines for the Penn Chinese Treebank (3.0)," *IRCS Report 00-07*, University of Pennsylvania, October 2000.
- [106] Nianwen Xue and Fei Xia, 2000. "The Bracketing Guidelines for the Penn Chinese Treebank (3.0)," *IRCS Report 00-08*, University of Pennsylvania, October 2000.

(7) Edited Volumes

- [107] Nancy Ide and Fei Xia (eds), 2012. Proceedings of the 6th Linguistic Annotation Workshop (The LAW VI), in conjunction with ACL 2012.
- [108] Fei Xia, William Lewis, and Lori Levin (eds), 2010. Proceedings of the ACL Workshop on NLP and Linguistics: Find the Common Ground, in conjunction with ACL/EACL 2010.
- [109] Chris Brew, Martha Palmer, and Fei Xia (eds), 2008. Proceedings of the 3rd Workshop on Issues in Teaching Computational Linguistics, in conjunction with ACL 2008.
- [110] Qing Ma and Fei Xia (eds), 2003. Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing (SIGHAN-2003), in conjunction with ACL 2003.
- [111] Martha Palmer, Mitch Marcus, Aravind Joshi, and Fei Xia (eds), 2000. Proceedings of the 2nd Chinese Language Processing Workshop (CLP-2000), in conjunction with ACL 2000.

(8) Manuscript:

- [112] Fei Xia, Imre Solti, and Ozlem Uzuner, 2009. "UW Internal Annotation Guidelines for the 2009 i2b2 Challenge and UW Medication IE System"
- [113] Ozlem Uzuner, Imre Solti, and Fei Xia, 2009. "i2b2 Medication Extraction Challenge Preliminary Annotation Guidelines"

[114] Ozlem Uzuner, Imre Solti, and Fei Xia, 2009. "i2b2 Medication Extraction Challenge Evaluation Metrics"

PRESENTATIONS

(1) Invited Talks

- [1] 2016 "Processing Clinical Data: Opportunities and Challenges", Data Science Workshop, Fred Hutchinson Cancer Research Center, Seattle, Jan 2016.
- [2] 2015 "Clinical NLP: Going beyond N-grams and Concepts", Oregon Health and Science University (OHSU), Portland, OR, March 2015.
- [3] 2014 "Applying NLP techniques to assist linguistic studies," University of British Columbia (UBC), Vancouver, Canada, Oct 2014.
- [4] 2014 "Rapid development of resources and NLP systems for resource-poor languages," IBM Almaden Research Center, San Jose, CA, Aug 2014.
- [5] 2014 "Domain adaptation via effective feature engineering across domains," UW/MS symposium, Redmond, WA, May 2014.
- [6] 2013 "Linguistics and natural language processing: A tale of two cities," Peking University, Beijing, China, Sept 2013.
- [7] 2013 "Clinical NLP Research: Challenges and Strategies," IBM T. J. Watson Research Center, Yorktown Heights, NY, April 2013.
- [8] 2012 "Clinical NLP Research: Challenges and Strategies," Microsoft Research Asia, Beijing, China, July 2012.
- [9] 2012 "Information Engineering and Synthesis for Resource-Poor Languages (RiPLes)," National Science of Academy, Beijing, China, July 2012.
- [10] 2010 "Collecting and Analyzing Language Data from the Web", UW/MS symposium, Redmond, WA, Jan 2010.
- [11] 2009 "Building a Large Multilingual Resource with Natural Language Processing Techniques," University of Victoria, Victoria, BC, Canada, Nov 27, 2009.
- [12] 2009 "Applying NLP Technologies to the Collection and Analysis of Language Data to Aid Linguistic Research," Simon Fraser University, Vancouver, Canada, Nov 26, 2009.
- [13] 2009 "A Two-step Approach to DS-to-PS Conversion," Workshop on building a Hindi/Urdu Treebank, Boulder, Colorado, June 7, 2009.
- [14] 2009 "Dependency structure and phrase structure," Workshop on building a Hindi/Urdu Treebank, Hyderabad, India, Jan 14, 2009.

- [15] 2008 “General techniques for creating treebanks”. NLP Winter School 2008. Hyderabad, India. Jan 6, 2008.
- [16] 2008 “Tapping Linguistic Data on the Web for Building Tools for Resource Poor Languages”. NLP Winter School 2008. Hyderabad, India, Jan 5, 2008. With William Lewis.
- [17] 2007 “Grammar Formalisms and Transformations between Formalisms”, a panelist on the Treebank Annotation Workshop, Rochester, NY, April 26, 2007.
- [18] 2007 “Harvesting and Enriching Multilingual Semi-structured Text to Facilitate Language Search”, Google, Mountain View, CA, Feb 20, 2007. With William Lewis.
- [19] 2006 “Towards automatic enrichment and analysis of linguistic data for low-density languages” UW/MS Symposium, Microsoft Research, Redmond, WA.
- [20] 2006 “Projecting Syntactic Structure using Interlinear Gloss Text”, Computational Linguistics Poster Session, University of Washington, Seattle, WA. With William Lewis and Dan Jinguji.
- [21] 2005 “Combining Linguistic Knowledge and Electronic Data for Natural Language Processing”
a. CMU, Pittsburg, PA.
b. University of Washington, Seattle, Washington.
- [22] 2002 “Syntax-Based Phrase Extraction for IBM MT System”, The 2002 MT Workshop, Santa Monica, CA.
- [23] 2001 “Automatic Grammar Generation from Two Different Perspectives”
c. University of Maryland, College Park, MD.
d. Microsoft Research, Redmond, WA.
e. IBM T. J. Watson Research Center, Yorktown Heights, NY.
- [24] 2001 “Relating Grammars and Treebanks for Natural Language Processing ”
f. AT&T Research, Florham Park, New Jersey
g. MIT Lincoln Lab, Lexington, MA.
h. CS Dept, Harvard University, Cambridge, MA
i. BBN Technology, Boston, MA.
- [25] 2001 “The Chinese Penn Treebank Project,” IBM T. J. Watson Research Center, Yorktown Heights, New York.
- [26] 2000 “A Uniform Method of Grammar Extraction and Its Application,” IBM NLP Open House, IBM T. J. Watson Research Center, Hawthorne, New York.
- [27] 2000 “Developing Guidelines and Ensuring Annotation Accuracy for the Chinese Penn Treebank” Computational Linguistics Series, CS Dept, Univ. of Maryland, College Park, Maryland.
- [28] 1999 “Towards Semi-automating Grammar Development”, AT&T Student Research Day, AT&T Research Lab, Florham Park, New Jersey.
- [29] 1999 “The first stage of the Chinese Penn Treebank Project”, At the Birds-of-a-feather session during ACL-1999, College Park, Maryland.

- [30] 1998 “Building a Chinese Treebank”, the International Conference on Chinese Information Processing (ICCIP-1998), Beijing, China.
- [31] 1998 “Grammar Development using Tree Descriptions”, Tutorial for the 4th TAG+ Workshop, Philadelphia, Pennsylvania.
- [32] 1998: “Chinese Tree Adjoining Grammar”, The 1st International Chinese Language Processing Workshop (CLP-1998), Philadelphia, Pennsylvania.

(2) Refereed Conference/Workshop Presentations

- [32] 2016. "A Web-framework for ODIN Annotation", demo at ACL 2016, Berlin, Germany. With Ryan Georgi and Michael Wayne Goodman.
- [33] 2016. “Annotating and Detecting Medical Events in Clinical Notes”, presented at LREC 2016, Portoroz, Slovenia. With Prescott Klassen and Meliha Yetisgen.
- [34] 2016. “Annotating and Detecting Medical Events in Clinical Notes”, presented at Northwest NLP (NW-NLP) Workshop, Seattle. With Prescott Klassen and Meliha Yetisgen.
- [34] 2015. “Extracting Topic-Specific Synonyms from Twitter”, poster at the 10th Annual Women in Machine Learning Workshop, Montreal, Canada. With Maria Antoniak and Eric Bell.
- [35] 2015. “Enriching Interlinear Text using Automatically Constructed Annotator”, at LaTeCH-2015, Beijing, China. With Ryan Georgi and William D. Lewis.
- [36] 2015. “Leveraging Paraphrase Labels to Extract Synonyms from Twitter”, in FLAIRS-2015, Hollywood, Florida, USA. With Maria Antoniak and Eric Bell.
- [37] 2015. “Enriching, Editing, and Representing Intelinear Glossed Text”, at CICLEing 2015, Cairo, Egypt. With Michael Wayne Goodman, Ryan Georgi, Glenn Slayden, and William D. Lewis.
- [38] 2014. “Unsupervised Dependency Parsing with Transferring Distribution via Parallel Guidance and Entropy Regularization”, poster at ACL-2014, Baltimore, MD. With Xuezhe Ma.
- [39] 2014. “Learning Grammar Specifications from IGT: A Case Study of Chintang”, at ComputEL workshop, Baltimore, MD. With Emily M. Bender, Joshua Crowgey, and Michael Wayne Goodman.
- [40] 2014. “Modern Chinese Helps Archaic Chinese Processing: Finding and Exploiting the Shared Properties”, at LREC 2014, Reykjavik, Iceland. With Yan Song.
- [41] 2014. “Enriching ODIN”, at LREC 2014, Reykjavik, Iceland. With William Lewis, Michael Wayne Goodman, Joshua Crowgey and Emily M. Bender.
- [42] 2014. “Annotating Clinical Events in Text Snippets for Phenotype Detection”, at LREC 2014, Reykjavik, Iceland. With Prescott Klassen, Lucy Vanderwende and Meliha Yetisgen.

- [43] 2014. "A New Corpus for Clinical Events with Change of State", at the 2014 AMIA Joint Summit on Translational Science, San Francisco. With Meliha Yetisgen-Yildiz, Prescott Klassen, and Lucy Vanderwende.
- [44] 2013. "Text Processing Tools from the University of Washington Biomedical Language Processing Group", in Proceedings of the 2013 AMIA Symposium. With Meliha Yetisgen-Yildiz, Cosmin A. Bejan, Prescott Klassen, Michael Tepper, and Lucy Vanderwende.
- [45] 2013. "The Construction of a Segmented and Part-of-speech Tagged Archaic Chinese Corpus: A Case Study on Huainanzi", at the 12th China National Conference on Computational Linguistics (CNCCL 2013), Suzhou, China. With Kam Tang Lau and Yan Song.
- [46] 2013. "A Common Case of Jekyll and Hyde: the Synergistic Effect of Using Divided Source Training Data for Feature Augmentation", at IJCNLP. Nagoya, Japan. With Yan Song.
- [47] 2013. "Towards Creating Precision Grammars from Interlinear Glossed Text: Inferring Large-scale Typological Properties", at the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2013), Sofia, Bulgaria. With Emily M. Bender, Michael Wayne Goodman, and Joshua Crowgey.
- [48] 2013. "Dependency Parser Adaptation with Subtrees from Auto-Parsed Target Domain Data", at ACL, Sofia, Bulgaria, Aug 2013. With Xuezhe Ma.
- [49] 2013. "Enhanced and Portable Dependency Projection Algorithms Using Interlinear Glossed Text", at ACL, Sofia, Bulgaria, Aug 2013. With Ryan Georgi and William D. Lewis.
- [50] 2013. "Modeling Annotator Rationales with Application to Pneumonia Classification", in Proceedings of the 2013 AAAI workshop on Expanding the Boundaries of Health Informatics Using Artificial Intelligence (HIAI 2013), July 15, Bellevue, WA. With Michael Tepper, Heather L. Evans, and Meliha Yetisgen-Yildiz.
- [51] 2013: "Towards Automatic Detection of Morphosyntactic Systems from IGT", at the Workshop on Exploring Data from Language Documentation. Berlin, Germany. With Emily M. Bender, Joshua Crowgey, and Michael Wayne Goodman.
- [52] 2013: "Automated Tools for Phenotype Extraction from Medical Records", poster at the 2013 AMIA Clinical Research Informatics Summit. With Meliha Yetisgen-Yildiz, Cosmin A. Bejan, Lucy Vanderwende, Heather L. Evans, and Mark M. Wurfel.
- [53] 2013: "Annotating Change of State for Clinical Events", poster at the 1st Workshop on Events: Definition, Detection, Coreference, and Representation, in conjunction with NAACL-2013, Atlanta, GA. With Lucy Vanderwende and Meliha Yetisgen-Yildiz.
- [54] 2012: "Improving Dependency Parsing with Interlinear Glossed Text and Syntactic Projection", Poster at COLING. Mumbai, India, Dec 2012. With Ryan Georgi and William D. Lewis.
- [55] 2012: "Entropy-based Training Data Selection for Domain Adaptation", short paper, In Proceedings of COLING. Mumbai, India, Dec 2012. With Yan Song and Prescott Klassen.
- [56] 2012. "Smoking Status Detection across Domains", Poster at the AMIA Fall Symposium, Chicago, Illinois, November 2012. With Michael Tepper and Meliha Yetisgen-Yildiz.

- [57] 2012: “Creating a Tree Adjoining Grammar from a Multilayer Treebank”, At the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11), Paris, France, September 2012. With Rajesh Bhatt and Owen Rambow.
- [58] 2012: “Effort of Genre Variation and Prediction of System Performance”, At LREC-2012, Istanbul, Turkey, May 22-25. With Dong Wang.
- [59] 2012: “Using a Goodness Measurement for Domain Adaptation: A Case Study on Chinese Word Segmentation”, At LREC-2012, Istanbul, Turkey, May 22-25. With Yan Song.
- [60] 2012: “Measuring the Divergence of Dependency Structures Cross-Linguistically to Improve Syntactic Projection Algorithms”, At LREC-2012, Istanbul, Turkey, May 22-25. With Ryan Georgi and William D. Lewis.
- [61] 2012: “Statistical Section Segmentation in Free-Text Clinical Records”. At LREC-2012, Istanbul, Turkey, May 22-25. With Michael Tepper, Daniel Capurro, Lucy Vanderwende, and Meliha Yetisgen-Yildiz
- [62] 2012: “Clinical corpus annotation: challenges and strategies”, At the third Workshop on Building and Evaluating Resources for Biomedical Text Mining, in conjunction with LREC-2012, Istanbul, Turkey. With Meliha Yetisgen-Yildiz.
- [63] 2012: “Challenges in Converting between Treebanks: a Case Study from the HUTB”, in Proceedings of META-RESEARCH Workshop on Advanced Treebanking, in conjunction with LREC-2012, Istanbul, Turkey. With Rajesh Bhatt.
- [64] 2011: “Linguistic Phenomena, Analyses, and Representations: Understanding Conversion between Treebanks”, At IJCNLP-2011, Chiang Mai, Thailand, Nov 9-13. With Rajesh Bhatt and Owen Rambow.
- [65] 2011: “Statistical Machine Translation for Biomedical Text: Are We There Yet?” At the AMIA Annual Symposium, Washington DC, Oct 22-26. With Cuijun Wu, Louise Deleger, and Imre Solti.
- [66] 2011. “Automatic Identification of Critical Follow-Up Recommendation Sentences in Radiology Reports”. At the AMIA Annual Symposium, Washington DC, Oct 22-26. With Meliha Yetisgen-Yildiz, Martin Gunn, and Tom Payne.
- [67] 2011. “Extraction of Pneumonia Cases from Free-Text Intensive Care Unit Reports”. At the AMIA 2011 Annual Symposium. Washington DC. Oct 22-26. With Meliha Yetisgen-Yildiz, Bradford Glavan, Lucy Vanderwende, and Mark Wurfel.
- [68] 2011. “Identifying Patients with Pneumonia from Free-Text Intensive Care Unit Reports”. At the ICML workshop on Learning from Unstructured Clinical Text, Bellevue, WA, July 2. With Meliha Yetisgen-Yildiz, Bradford Glavan, Lucy Vanderwende, and Mark Wurfel.
- [69] 2011. “Email Formality in the Workplace: A Case Study on the Enron Corpus,” At the 2011 ACL Workshop on Language in Social Media (LSM 2011), Portland, Oregon, June 23. With Kelly Peterson and Matt Hohensee.

- [70] 2010. "Addressing the Annotation Bottleneck for Clinical Natural Language Processing: Testing the Feasibility of Domain Adaptation for Medical Text," At the AMIA 2010 Annual Symposium, Washington DC, Nov 13-17. With Imre Solti and Scott Halgrim.
- [71] 2010. "Using Amazon's Mechanical Turk for Annotating Medical Named Entities," At the AMIA 2010 Annual Symposium, Washington DC, Nov 13-17. With Meliha Yetisgen-Yildiz and Imre Solti.
- [72] 2010. "Comparing Language Similarity Across Genetic and Typologically-Based Groupings," At COLING 2010, Beijing, China, Aug 23-27. With Ryan Georgi and William Lewis.
- [73] 2010. "A comparison of unsupervised methods for Part of Speech Tagging in Chinese," At COLING 2010, Beijing, China, Aug 23-27. With Alex Cheng and Jianfeng Gao.
- [74] 2010. "Extracting medication information from discharge summaries," At the Second Louhi Workshop on Text and Data Mining of Health Documents, in conjunction with NAACL 2010, Los Angeles, CA, June 5. With Scott Halgrim and others.
- [75] 2010. "Preliminary Experience with Amazon's Mechanical Turk for Annotating Medical Named Entities," At the NAACL workshop on Creating Speech and Text Language Data with Amazon's Mechanical Turk, Los Angeles, CA, June 6. With Meliha Yetisgen-Yildiz and others.
- [76] 2010. "Empty Categories in a Hindi Treebank," At the 7th International Conference on Language Resources and Evaluation (LREC-2010), Malta, May 17-23, 2010. With Archana Bhatia and others.
- [77] 2010. "The Problems of Language Identification within Hugely Multilingual Data Sets," At the 7th International Conference on Language Resources and Evaluation (LREC-2010), Malta, May 17-23, 2010. With Carrie Lewis and Will Lewis.
- [78] 2010. "Peeling Away the Black Box Label: Clinical Validation of a MaxEnt Machine Learning Character N-gram Feature Set for Acute Lung Injury," At the 2010 AMIA Summit on Translational Bioinformatics, San Francisco, CA, March 10-12, 2010. With Imre Solti and others.
- [79] 2010. "Statistical Extraction of Medication Information from Clinical Records," At the 2010 AMIA Summit on Translational Bioinformatics, San Francisco, CA, March 10-12, 2010. With Scott Russell Halgrim and others.
- [80] 2010. "*Language ID for a Thousand Languages*," At LSA-2010, Baltimore, Maryland, Jan 7-10, 2010. With Carrie Lewis and William Lewis.
- [81] 2009. "Hindi Syntax: Annotating Dependency, Lexical Predicate-Argument Structure, and Phrase Structure", At the 7th International Conference on Natural Language Processing (ICON-2009), Hyderabad, India, Dec 14-17, 2009. With Martha Palmer and others.
- [82] 2009. "Automated classification of radiology reports for acute lung injury: Comparison of keyword and machine learning based natural language processing approaches," At the *NLP Workshop, IEEE International Conference on Bioinformatics and Biomedicine (BIBM-2009)*, Washington DC, November 1-4, 2009. With Imre Solti and others.
- [83] 2009. "A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu," At the Third Linguistic Annotation Workshop (The LAW III) in conjunction with ACL/IJCNLP 2009. Singapore. Aug 6-7, 2009. With Rajesh Bhatt and others.

- [84] 2009 “Language ID in the Context of Harvesting Language Data off the Web,” At the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2009), Athens, Greece, March 30 – April 3, 2009. With William Lewis and Hoifung Poon.
- [85] 2009 “Parsing, Projecting & Prototypes: Repurposing Linguistic Data on the Web,” At the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2009), Demo session, Athens, Greece, March 30 – April 3, 2009. With William Lewis.
- [86] 2009 “Applying NLP Technologies to the Collection and Enrichment of Language Data on the Web to Aid Linguistic Research,” At the workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH-SHELT&R 2009). In conjunction with EACL-2009, Athens, Greece, March 30, 2009. With William Lewis.
- [87] 2009 “Towards a Multi-Representational Treebank,” At the 7th International Workshop on Treebanks and Linguistic Theories (TLT 2009), Groningen, Netherlands, Jan 23-24, 2009. With Owen Rambow and others.
- [88] 2008 “The evolution of a statistical NLP course,” At the 3rd *ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, Columbus, Ohio, June 19-20, 2008.
- [89] 2008 “Building a flexible, collaborative, intensive master’s program in computational linguistics,” At the 3rd *ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL*, Columbus, Ohio, June 19-20, 2008. With Emily Bender and Erik Bansleben.
- [90] 2008 “Repurposing Theoretical Linguistic Data for Tool Development and Search,” At *IJCNLP-2008*, Hyderabad, India. With William Lewis.
- [91] 2008 “Automatically Identifying Computationally Relevant Typological Features,” At *IJCNLP-2008*, Hyderabad, India. With William Lewis.
- [92] 2008 “A Hybrid Approach to the Induction of Underlying Morphology,” At *IJCNLP-2008*. With Michael Pepper.
- [93] 2008 “Finding parallel texts on the web using cross- language information retrieval”, At the *Workshop on Cross Language Information Access* in conjunction with *IJCNLP-2008*. With Achim Ruopp.
- [94] 2007 “Multilingual Structural Projection across Interlinearized Text, ” At *NAACL-HLT 2007*, Rochester, New York. With William Lewis.
- [95] 2007 “Projecting Structure onto Data for Resource-Poor and Endangered Languages”, At the 81st Annual Meeting of Linguistic Society of America (LSA-2007), Anaheim, CA. With William Lewis and Dan Jinguji.
- [96] 2006 “Enriching Language Data through Projected Structures”, At *TLSX*, Austin, Texas. With William Lewis and Dan Jinguji.
- [97] 2006 “Features, Bagging, and System Combination for the Chinese POS Tagging Task,” At *SIGHAN 2006*, Sydney, Australia. With Lap Cheung.

- [98] 2004 “Improving a Statistical MT System with Automatically Learned Rewrite Patterns,” At COLING-2004, Geneva, Switzerland. With Michael McCord.
- [99] 2003 “A Phrase-Based Unigram Model for Statistical Machine Translation,” At HLT/NAACL-2003, Edmonton, Canada. With Christoph Tillmann.
- [100] 2003 “TIPS: A Translingual Information Processing System,” At HLT-NAACL 2003, Demo Session, Edmonton, Canada. With Yaser Al-Onaizan and others.
- [101] 2001 “Automatically Extracting and Comparing Lexicalized Grammars for Different Languages,” At IJCAI-2001, Seattle, Washington. With Chung-hye Han, Martha Palmer and Aravind Joshi.
- [102] 2001 “Converting Dependency Structures to Phrase Structures,” At HLT-2001, San Diego, CA. With Martha Palmer.
- [103] 2000 “A Uniform Method of Grammar Extraction and Its Applications,” At EMNLP/VLC-2000, Hong Kong. With Martha Palmer, and Aravind Joshi.
- [104] 2000 “Developing Guidelines and Ensuring Consistency for Chinese Text Annotation,” At LREC-2000, Athens, Greece. With Martha Palmer, Nianwen Xue, Mary Ellen Okurowski, John Kovarik, Fu-Dong Chiou, Shizhe Huang, Tony Kroch, and Mitch Marcus.
- [105] 2000 “Comparing Lexicalized Treebank Grammars Extracted from Chinese, Korean, and English Corpora,” At CLP-2000, Hong Kong, October 8, 2000. With Chung-hye Han, Martha Palmer, and Aravind Joshi.
- [106] 2000 “Some Experiments on Indicators of Parsing Complexity for Lexicalized Grammars,” At Workshop on Efficiency in Large-Scale Parsing Systems Workshop, Luxembourg, Germany. With Anoop Sarkar and Aravind Joshi.
- [107] 2000 “Evaluating the Coverage of LTAGs on Annotated Corpora,” At the Workshop on Using Evaluation within HLT Programs: Results and Trends, Athens, Greece. With Martha Palmer.
- [108] 2000 “A Corpus-based Evaluation of Syntactic Locality in TAGs,” At TAG+ 2000, Paris, France. With Tonia Bleam.
- [109] 2000 “Comparing and Integrating Tree Adjoining Grammars,” At TAG+ 2000, Paris, France. With Martha Palmer.
- [110] 1999 “Extracting Tree Adjoining Grammars from Bracketed Corpora,” At NLPRS-1999, Beijing, China.
- [111] 1999 “Towards Semi-automating Grammar Development,” At NLPRS-1999, Beijing, China. With Martha Palmer, and K. Vijay-Shanker.
- [112] 1998 “Consistent Grammar Development Using Partial-Tree Descriptions for LTAGs,” at TAG+ 1998, Philadelphia. With Martha Palmer, K. Vijay-Shanker, and Joseph Rosenzweig.
- [113] 1997 “A Comparison of Head Transducers and Transfer for a Limited Domain Translation,” at ACL-1997, Madrid, Spain. With Hiyaw Alshawi, Adam Buchsbaum.

- [114] 1997 “English-to-Mandarin Speech Translation with Head Transducers,” At the Workshop of Spoken Language Translation (SLT-1997), Madrid, Spain, July 11, 1997. With Hiyan Alshawi.
- [115] 1997 “Maintaining the Forest and Burning out the Underbrush in XTAG,” At the Workshop on Computational Environments for Grammar Development and Language Engineering (ENVGRAM-1997), Madrid, Spain. With Christine Doran, Beth Hockey, Philip Hopely, Joseph Rosenzweig, Anoop Sarkar, Srinivas Bangalore, Alexis Nasr and Owen Rambow.
- [116] 1996 “Capturing Language Specific Constraints on Lexical Selection with Feature-Based LTAGs,” At the International Conference on Chinese Computing (ICCC-1996), Singapore, June 4 – 7, 1996. With Chung-hye Han, Martha Palmer, and Joseph Rosenzweig.

PROFESSIONAL ACTIVITIES

(1) Conference and Workshop Organization

- [1] Area Chair, NAACL 2016, San Diego, CA, June 12-17, 2016.
- [2] Publicity Co-chair, the 3rd CCF Conference on Natural Language Processing and Chinese Computing (CCF-NLPCC), Shenzhen, China, Dec 2014.
- [3] Tutorial Co-chair, COLING, Dublin, Ireland, Aug 2014.
- [4] Program Co-chair, the 2nd CIPS-SIGHAN Joint Conference on Chinese Language Processing, Tianjin, China, Dec 20-21, 2012.
- [5] Co-chair, the 6th Linguistic Annotation Workshop (LAW) in conjunction with ACL-2012, Jeju, Republic of Korea, July 12-13, 2012.
- [6] Area Chair, ACL 2012, Jeju, Republic of Korean, July 8-14, 2012.
- [7] Co-chair, the workshop on “Using Treebanks for Linguistic Research: the View from the Hindi-Urdu Treebank” at the 2011 Linguistic Summer Institute, July 23, 2011, Boulder, Colorado.
- [8] Co-chair, the workshop on “NLP and Linguistics: finding the common ground”. In conjunction with ACL-2010, Uppsala, Sweden, July 15-16, 2010.
- [9] Co-chair, the Third Workshop on Issues in Teaching Computational Linguistics. In conjunction with ACL-2008, June 2008.
- [10] Exhibition/Demo Chair, IJCNLP-08, Hyderabad, India, January 8-10, 2008.
- [11] Co-chair, the Workshop on Treebank Annotation, Rochester, NY, April 26, 2007.
- [12] Senior Program Committee, Human Language Technology Conference and North American Chapter of ACL (HLT/NAACL 2006), New York City, June 5-7, 2006.

- [13] Co-chair, the 2nd SIGHAN Workshop on Chinese Language Processing (SIGHAN-2003) in conjunction with ACL-2003, Sapporo, Japan, July 11 – 12, 2003.
- [14] Co-chair, the 2nd International Chinese Language Processing Workshop (CLP-2000), Hong Kong, October 8, 2000.
- [15] Co-chair, the 1st International Chinese Language Processing Workshop (CLP-1998), Philadelphia, PA, June 30 – July 2, 1998.

(2) Conference and Workshop Reviewing

- [16] ACL (Association of Computational Linguistics): 2016, 2015, 2014, 2012, 2010, 2008, 2007, 2006, 2002.
- [17] HLT/NAACL (North American Chapter of ACL): 2015, 2014, 2008, 2007, 2006, 2005
- [18] EACL (European Chapter of ACL): 2014, 2009
- [19] COLING (International Conference on Computational Linguistics): 2016, 2012, 2010, 2006, 2004, 2002
- [20] EMNLP (Empirical Methods in Natural Language Processing): 2010, 2009, 2008, 2007, 2005
- [21] IJCNLP (International Joint Conference on Natural Language Processing): 2013
- [22] LREC (Language and Resource Evaluation Conference): 2016, 2014
- [23] AMIA (American Medical Informatics Association): 2016, 2013, 2012, 2011.
- [24] IJCAI (International Joint Conference on AI): 2013.
- [25] AMTA (Biennial Conference of the Association for Machine Translation in the Americas): 2008, 2006
- [26] ICON (International Conference on Natural Language Processing): 2015, 2010, 2009
- [27] PACLIC (Pacific Asia Conference on Language, Information, and Computation): 2010, 2009
- [28] SIGHAN Workshop on Chinese Language Processing: 2010, 2006, 2002
- [29] LAW (Linguistics Annotation Workshop): 2016, 2015, 2014, 2013
- [30] IEEE NLP-KE (IEEE International Conference on Natural Language Processing and Knowledge Engineering): 2005, 2003, 2001
- [31] TAG+ (International Workshop on Tree Adjoining Grammar and Related Formalisms): 2012, 2010, 2004, 1998.

- [32] Corpus Linguistics Fest (CLiF): 2016
- [33] SEM 2014 (the 3rd Joint Conference on Lexical and Computational Semantics)
- [34] NW-NLP (Pacific Northwest Regional NLP Workshop): 2014, 2012, 2010
- [35] ComputEL (Workshop on the Use of Computational Methods in the Study of Endangered Languages): 2014
- [36] CNCCL (China National Conference on Computational Linguistics): 2013
- [37] ESSLLI-2013 Workshop on Extrinsic Parse Improvement (EPI), 2013.
- [38] Workshop on Machine Translation and Parsing in Indian Languages (MTPIL), in conjunction with COLING 2012, Mumbai, India, Dec 15-16, 2012.
- [39] META-RESEARCH Workshop on Advanced Treebanking, in conjunction with LREC-2012, Istanbul, Turkey, May 22, 2012.
- [40] Inconsistency Robustness, Stanford University, Aug 16-18, 2011.
- [41] Workshop on Issues on Teaching Computational Linguistics, in conjunction with ACL: 2008
- [42] Deep Processing Workshop in conjunction with ACL-07, Prague, Czech, June 2007.
- [43] The 21st International Conference on the Computer Processing of Oriental Languages (ICCPOL 2006), Singapore, Dec 17-19, 2006.
- [44] LREC Workshop on Merging and Layering Linguistic Information, Genoa, Italy, May 23, 2006.
- [45] Frontiers in Corpus Annotation 2004 NACCL/HLT Conference Workshop, Boston, MA, May 6, 2004.
- [46] The 2001 International Conference on Chinese Computing (ICCI-2001), Singapore, November 27 – 29, 2001.

(3) Journal and Book Reviewing

- [47] Editorial Board Member, Journal of Computational Linguistics (CL), 2014-now.
- [48] Editorial Board Member, Text, Speech, and Language Technology book series, 2012-present
- [49] Editorial Board Member, Journal of Language Resources and Evaluation (LRE), 2010-now.

- [50] Associate Editor, Journal of ACM Transactions on Asian Language Information Processing (TALIP), 2007-2010.
- [51] Reviewer, The Handbook of Linguistics Annotation (edited by Nancy Ide and James Pustejovsky), Springer Press
- [52] Reviewer, Journal of American Medical Informatics Association (JAMIA), 2011
- [53] Reviewer, Journal of Language Resources and Evaluation (LRE), 2016, 2015, 2014, 2005-2010.
- [54] Reviewer, Journal of ACM Transactions on Asian Language Information Processing (TALIP), 2014, 2003-2007.
- [55] Reviewer, Journal of Computational Linguistics (CL), 2015, 2011, 2010, 2009, 2008, 2006, 2005
- [56] Reviewer, Journal of Information Technology and Decision Making (IJITDM), 2009.
- [57] Reviewer, Journal of IEEE Transactions on Audio, Speech and Language Processing, 2008.
- [58] Reviewer, Journal of IEEE Intelligent Systems, 2008
- [59] Reviewer, Journal of Computer Science and Technology (JCST), 2007.
- [60] Reviewer, Journal of Computational Linguistics and Chinese Language Processing (CLCLP), 2007.
- [61] Reviewer, *Journal of Computational Linguistics and Chinese Language Processing (CLCLP)*, 2003.
- [62] Reviewer, *Journal of AI Communications (AICOM): the European Journal on Artificial Intelligence*, 2001.
- [63] Reviewer, *CSLI Publication on Collaborative Language Engineering*, 2001.

(4) Other

- [64] Member of the Executive Board, the North American Chapter of the Association for Computational Linguistics (NAACL), 2015-now.
- [65] NSF proposal review, 2016, 2014, 2013, 2010, 2008.
- [66] Reviewer, Research Grants Council (RGC) of Hong Kong, 2015, 2014, 2013, 2003 – 2005.
- [67] Officer, SIGANN (ACL Special Interest Group for Annotation), 2012-now
- [68] Officer, SIGHAN (ACL Special Interest Group on Chinese Language Processing), 2012-2013, 2002-2003.

UW COMMITTEES AND OTHER DUTIES

(1) Linguistics Committees:

- [1] A member of the resource committee, 2012-present.
- [2] Co-organizer of UW/Microsoft Symposium, 2007-present.
- [3] A member of the CLMS Program Admission Committee, 2005-present.
- [4] Co-organizer of the CLMS Orientation, 2005-present.
- [5] A member of Lance Forshay's promotion review committee, fall 2012.
- [6] Acting Director of Computational Linguistics Lab: Fall 2005, 12/16/2010-6/15/2011.
- [7] A committee member for the Computational Linguistics faculty search in the UW Linguistics Department, 2009-2010.
- [8] Team leader of the UW team for the i2b2 contest, 2009
- [9] A member of the Graduate Research Excellence Award Committee: 2007, 2010.
- [10] A member of the ad-hoc committee on MA and Ph.D. requirement, Spring 2006.

(2) UW committee:

- [11] The Scientific Review Panel for the Institute of Translational Health Sciences (ITHS) Small Pilot and Technology & Resources Access grant programs, 2010.
- [12] The Review Committee member for the NLP faculty search in Dept of Medical Education and Biomedical Informatics, UW School of Medicine, 2009.
- [13] A member of the Task Force on Digital Humanities, School of Arts and Sciences, 2007-2008. The work led to the successful application of an NEH Digital Humanities Challenge Grant by the Simpson Center.
- [14] Reviewer for the UW Royal Research Fund (RRF): 2014, 2013, 2010, 2007.

HONORS AND AWARDS

06/2015 – 06/2016 IBM Faculty Award

09/2000 – 09/2001 IRCS Fellowship, University of Pennsylvania

09/1995 – 08/2000 Research Fellowship, University of Pennsylvania
09/1994 – 08/1995 Dean's Fellowship, University of Pennsylvania
09/1991 – 07/1992 Lenovo Awards, Peking University
09/1990 – 07/1991 Outstanding Student Award, Peking University
09/1989 – 07/1990 Outstanding Student Award, Peking University

TEACHING EXPERIENCE

09/2005-present University of Washington
Shallow Processing Techniques for NLP: 2007, 2009, 2010, 2013-2016
Deep Processing Techniques for NLP: 2005
Advanced Statistical Methods in NLP: 2006-2011, 2013-2017
Seminar on Machine Translation: 2006, 2007
Seminar on Unsupervised Learning: 2008
Seminar on Bridging Linguistics and NLP: 2009
Seminar on Information Extraction from Biomedical text: 2010
Seminar on Computational Linguistics and Social Media: 2011
Seminar on Domain Adaptation: 2013
Seminar on Creating Resources for Resource-poor Languages: 2014, 2016
Seminar on Declarative Information Extraction: 2015

07/07/2011-08/02/2011: Teach at the 2011 Linguistic Summer Institute, Boulder, Colorado.
“Treebanking and Theoretical Linguist”

GRADUATE STUDENTS SUPERVISED

(1) Doctoral Committees

A. Chair of Committee

- [1] Prescott Klassen: completed in summer 2016
 - Dissertation title: “Defining, Extracting, and Applying Events in NLP Tasks for Clinical Corpora”

- [2] Ryan Georgi: Spring 2009-2016, completed in summer 2016
 - Dissertation title: “From Aari to Zulu: Massively Multilingual Creation of Language Tools Using Interlinear Glossed Text”

B. Member of the Reading Committee

- [1] Chad Mills: Linguistics
- [2] Michael Goodman: Linguistics
- [3] Wen-Wai Yim: Biomedical and Health Informatics, completed in spring 2016
 - Dissertation title: “Information Extraction from Clinical and Radiology Notes for Liver Cancer Staging”
- [4] Amittai Axelrod: Electronic Engineering, Completed in summer 2014
 - Dissertation title: “Data Selection for Statistical Machine Translation”
- [5] Sanghoun Song: Linguistics, Completed in 12/2013
 - Dissertation title: “A Grammar Library for Information Structure”
- [6] Jeremy Kahn: Linguistics, Completed in 12/2010
 - Dissertation title: “Parse decoration of the word sequence in the speech-to-text machine-translation pipeline”
- [7] Dustin Hillard: Electronic Engineering, Completed in 01/2008.
 - Dissertation title: “Automatic Sentence Structure Annotation for Spoken Language Processing”
- [8] Meliha Yetisgen-Yildiz: Information School, Completed in Fall 2007.
 - Dissertation title: “Using Statistical and Knowledge Based Approaches for Literature-Based Discovery”

C. Member, Graduate School Representative (GSR)

- [9] Tony Fader: Computer Science and Engineering, completed in spring 2014.
- [10] Guangbin Zhuang: Mathematics, completed in summer 2013.
- [11] Congle Zhang: Computer Science and Engineering, completed in summer 2013.
- [12] Zixia Hu: Electronic Engineering, completed in fall 2012
- [13] Raphael Domink Hoffmann: Computer Science and Engineering, completed in summer 2012
- [14] Stanley Kok: Computer Science and Engineering, Completed in Spring 2010

D. Member of PhD committee outside UW

- [15] Rafiya Begum, International Institute of Information Technology (IIIT), Hyderabad, India.
- [16] Joel Dunham, University of British Columbia, Vancouver, Canada, completed in fall 2014.
 - Dissertation title: “The Online Linguistic Database: Software for linguistic fieldwork”

- [17] Yan Song, City University of Hong Kong, Hong Kong, China.
➤ Dissertation title: “Transductive learning, data refactoring, and model reweighting in domain adaptation for language processing”
- [18] Daniel Gar-Shon Tse, University of Sydney, Sydney, Australia, completed in summer 2013.
➤ Dissertation title: “Chinese CCGBank: Deep derivations and dependencies for Chinese CCG parsing”

(2) Master Committees

A. Adviser, master’s thesis

- [1] Faraz Waseem
- [2] Ronald Lockwood: completed in spring 2015
➤ Thesis: “A Linguist-Friendly Machine Translation System for Low-Resource Languages”
- [3] Maria Antoniak: completed in fall 2014
➤ Thesis: “Extracting Topically Related Synonyms from Twitter using Syntactic and Paraphrase Data”
- [4] Claire Jaja: completed in fall 2014
➤ Thesis: “Leveraging Training Data from High-Resource Languages to Improve Dependency Parsing for Low-Resource Languages”
- [5] Prescott Klassen: completed in winter 2012
➤ Thesis: “Calculating LLR Topic Signatures with Dependency Relations for Automatic Text Summarization”
- [6] Nicholas Flacco: Completed in spring 2011
➤ Thesis: “Using Semantic Information to Improve Protein Interaction Detection”
- [7] Eric Bell: Completed in summer 2010
➤ Thesis: “Event Detection and Classification by Sense Disambiguation”
- [8] Alex Cheng: Completed in winter 2010
➤ Thesis: “Survey of unsupervised methods for Chinese POS tagging with Hidden Markov Models”
- [9] Scott Russell Halgrim: Completed in fall 2009.
➤ Thesis: “A Pipeline Machine Learning Approach to Biomedical Information Extraction”
- [10] Ryan Georgi: Completed in winter 2009

- Thesis: “PCFG Induction using Interlinear-derived Prototypes”
- [11] Ankit Srivastava: Completed in spring 2008
 - Thesis: “Learning a Translation Lexicon from Non-parallel Corpora”
- [12] Michael Tepper: Completed in fall 2007
 - Thesis: “Knowledge-Lite Induction of Underlying Morphology: Hybrid Approach to Learning Morphemes Using Context-Sensitive Rewrite Rules”
- [13] David Arthur Bullock: Completed in fall 2006,
 - Thesis: “TreeTran: A Tool for Visual Selection a Testing of Transfer Rules for Machine Translation”
- [14] Achim Ruopp: Completed in fall 2006
 - Thesis: “Finding and Evaluating Structured Bilingual Corpora on the Web”

B. Reader, MA thesis

- [15] Michael Lockwood: completed in winter 2016
 - Thesis: “Automated Gloss Mapping for Inferring Grammatical Properties”
- [16] Andrea Kahn: completed in summer 2015
 - Thesis: “New Methods for Detecting Deceptive Product Reviews”
- [17] David Wax: completed in winter 2014
 - Thesis: “Automated Grammar Engineering for Verbal Morphology”
- [18] Todd Lingren: completed in summer 2013
 - Thesis: “Detecting Adverse Events in Clinical Trial Free Text”
- [19] Chad Mills: completed in spring 2013
 - Thesis: “Learning Board Game Rules from an Instruction Manual”
- [20] Sergei Lushtak: completed in fall 2012
 - Thesis: “Unsupervised Morphological Word Clustering”
- [21] Matt Hohensee: completed in winter 2012
 - Thesis: “It’s Only Morpho-Logical: Modeling Agreement in Cross-linguistic Dependency Parsing”
- [22] Spencer Rarrick: completed in fall 2011
 - Thesis: “Machine Translation Detection and Automatic Filtering of Web-Extracted Parallel Corpora”
- [23] Jason Shaw: completed in summer 2011
 - Thesis: “Learning for Resource-Poor Languages: Building a Language-Independent Model for Frame-Semantic Annotation”
- [24] Liam McGrath: completed in winter 2011

- Thesis: “Semantic Features for Semantic Role Labeling”
- [25] Safiyyah Saleem, completed in spring 2010
 - Thesis: “Augment Optionality: A New Library for the Grammar Matrix Customization System”
- [26] Eric Mead, Completed in spring 2010
 - Thesis: “Associative Anaphora Resolution Using Word Sense Disambiguation and Spreading Activation in an Augmented Ontology”
- [27] David Potter, Simon Fraser University, completed in fall 2009
 - Thesis: “A Multiple Dominance Analysis of Sharing Coordination Constructions Using Tree Adjoining Grammar”
- [28] Kelly O’Hara: Completed in fall 2008
 - Thesis: “A Morphotactic Infrastructure for a Grammar Customization System”
- [29] Ping Yu: Completed in winter 2007:
 - Thesis: “GMM-based Automatic Identification between Mainland Mandarin and Taiwan Mandarin”
- [30] Albert A. Bertram: Completed in fall 2006:
 - Thesis: “An Application of a Connectionist Cognitive Model to Word Sense Disambiguation”

C. Advisor, for CLMA students who take the internship/report option

- [31] Chris Laterza
- [32] Jason Blum
- [33] John Cadigan
- [34] Micaela Tolliver
- [35] Ryan Martin
- [36] Wei Dai
- [37] William Lane
- [38] Joshua Valdez

- [39] Kenedy Anderson, completed in summer 2016
- [40] Alex Burrell, completed in summer 2016
- [41] Elizabeth Cary, completed in summer 2016
- [42] Jeff Health, completed in spring 2016
- [43] Anca Burducea, completed in fall 2015
- [44] Andrew Baer, completed in fall 2015
- [45] John Ho, completed in fall 2015

- [46] Ryan Bielby, completed in summer 2015
- [47] Katherine Vedder, completed in summer 2015
- [48] Rebecca Myhre, completed in summer 2015
- [49] Michael Foster, completed in spring 2015
- [50] Matthew Benak, completed in winter 2015
- [51] Ahmad Elshenawy, completed in winter 2015

- [52] Laurel Hart, completed in summer 2014
- [53] Alena Hrynkevich, completed in summer 2014
- [54] Yi-Shu Wei, completed in summer 2014
- [55] Jared Kramer, completed in spring 2014
- [56] Varden Wang, completed in winter 2014
- [57] David Lin, complete in fall 2013

- [58] Chase Hermsen, completed in summer 2013
- [59] Lisa Gress, completed in summer 2013
- [60] Kathryn Nichols, completed in summer 2013
- [61] Tristan Bodding-Long, completed in spring 2013
- [62] Edward Pham, completed in fall 2012

- [63] Greg Hullender, completed in spring 2012
- [64] Abdullah Alotayq, completed in winter 2012
- [65] Dong Wang, completed in fall 2011
- [66] Ryan Cross, completed in fall 2011
- [67] Jeffrey Shu, completed in fall 2011
- [68] Anthony Wong, completed in fall 2011
- [69] Kelly Peterson, completed in fall 2011

- [70] Imre Solti, completed in spring 2011
- [71] Matt Todd, completed in winter 2011
- [72] John Keesling, completed in winter 2011
- [73] Miles Pederson, completed in winter 2011
- [74] Parinta Parinta, completed in fall 2010

- [75] Ashley Gill, graduated in summer 2010
- [76] Michelle Stephen: graduated in spring 2010
- [77] Eduardo Alvarez-Godinez: graduated in winter 2010

- [78] Geoff Thilo: graduated in summer 2009
- [79] Eric Larsson, graduated in summer 2009
- [80] David Landan: graduated in spring 2009
- [81] Bob New: graduated in winter 2009

- [82] Nolan Lawson: Graduated in summer 2008
- [83] Sinclair Lindsey: Graduated in summer 2008
- [84] Dawei Hou: Graduated in spring 2008
- [85] Yohei Sakata: Graduated in fall 2007
- [86] Sabrina Burleigh: Graduated in fall 2007

- [87] Mike O'Leary: Graduated in summer 2007
- [88] Kathleen Sickles: Graduated in summer 2007
- [89] Jeffrey Karres: Graduated in summer 2007
- [90] Joshua Johanson: Graduated in spring 2007

MEMBERSHIPS

Association for Computational Linguistics
Linguistic Society of America

Last updated on 10/09/2016