

Department of Computer Science
University of Rochester
PO Box 270226
Rochester NY 14627

gildea@cs.rochester.edu
<http://www.cs.rochester.edu/~gildea/>
585 275-7230

EDUCATION

- PhD 2001 Computer Science, University of California, Berkeley.
Directed by Nelson Morgan, Daniel Jurafsky, Charles Fillmore, Jerome Feldman.
Dissertation: *Statistical Language Understanding Using Frame Semantics*
- MS 1999 Computer Science, University of California, Berkeley.
Thesis: *Topic-Based Language Models Using EM*
- BA 1995 Double major in Linguistics and Computer Science,
University of California, Berkeley.

RESEARCH INTERESTS

- statistical machine translation
- natural language understanding, including syntactic parsing and semantic interpretation
- machine learning algorithms

PROFESSIONAL EXPERIENCE

- 7/09 – Associate professor, Department of Computer Science, University of Rochester.
- 9/03 – 6/09 Assistant professor, Department of Computer Science, University of Rochester.
- 9/01 – 8/03 Postdoctoral fellow, Institute for Research in Cognitive Science and Department of Computer and Information Science, University of Pennsylvania.
- 8/96 – 8/01 Research assistant, International Computer Science Institute, Berkeley, California.
- 7/95 – 3/96 Software engineer, PA-RISC compiler group, Hewlett-Packard Corporation.

PUBLICATIONS

Books

Semantic Role Labeling Martha Palmer, Daniel Gildea, and Nianwen Xue. Synthesis Lectures on Human Language Technology Series. Morgan and Claypool, 2010.

Journal Articles

Grammar factorization by tree decomposition

Daniel Gildea. *Computational Linguistics*, 37(1):231–248, 2011.

Do grammars minimize dependency length?

Daniel Gildea and David Temperley. *Cognitive Science*, 34(2):286–310, 2010.

Binarization of synchronous context-free grammars

Liang Huang, Hao Zhang, Daniel Gildea, and Kevin Knight. *Computational Linguistics*, 35(4):559–595, 2009.

Enumeration of factorizable multi-dimensional permutations

Hao Zhang and Daniel Gildea. *Journal of Integer Sequences*, 07(5.8), 2007.

The Proposition Bank: An annotated corpus of semantic roles

Martha Palmer, Daniel Gildea, and Paul Kingsbury. *Computational Linguistics*, 31(1):71–106, 2005.

Effects of disfluencies, predictability, and utterance position on word form variation in English conversation

Alan Bell, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, Michelle Gregory, and Daniel Gildea. *Journal of the Acoustical Society of America*, 113(2):1001–1024, 2003.

Automatic labeling of semantic roles

Daniel Gildea and Daniel Jurafsky. *Computational Linguistics*, 28(3):245–288, 2002.

Learning bias and phonological rule induction

Daniel Gildea and Daniel Jurafsky. *Computational Linguistics*, 22(4):497–530, 1996.

Refereed Conference Papers

Tuning as linear regression

Marzieh Bazrafshan, Tagyoung Chung, and Daniel Gildea. In *North American chapter of the Association for Computational Linguistics (NAACL) short paper*, 2012.

Optimal head-driven parsing complexity for linear context-free rewriting systems

Pierluigi Crescenzi, Daniel Gildea, Andrea Marino, Gianluca Rossi, and Giorgio Satta. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics (ACL-11)*, 2011.

Issues concerning decoding with synchronous context-free grammar

Tagyoung Chung, Licheng Fang, and Daniel Gildea. In *Proceedings of the ACL 2011 Conference Short Papers*, Portland, Oregon, 2011.

Terminal-aware synchronous binarization

Licheng Fang, Tagyoung Chung, and Daniel Gildea. In *Proceedings of the ACL 2011 Conference Short Papers*, Portland, Oregon, 2011.

Effects of empty categories on machine translation

- Tagyoung Chung and Daniel Gildea. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-10)*, Boston, 2010.
- A fast fertility hidden markov model for word alignment using mcmc
Shaojun Zhao and Daniel Gildea. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-10)*, Boston, 2010.
- Semantic role features for machine translation
Ding Liu and Daniel Gildea. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING-10)*, Beijing, 2010.
- Optimal parsing strategies for Linear Context-Free Rewriting Systems
Daniel Gildea. In *Proceedings of the 2010 Meeting of the North American chapter of the Association for Computational Linguistics (NAACL-10)*, pages 769–776, 2010.
- Weight pushing and binarization for fixed-grammar parsing
Matt Post and Daniel Gildea. In *International Conference on Parsing Technologies*, Paris, France, October 2009.
- Bayesian learning of phrasal tree-to-string templates
Ding Liu and Daniel Gildea. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*, Singapore, 2009.
- Unsupervised tokenization for machine translation
Tagyoung Chung and Daniel Gildea. In *Conference on Empirical Methods in Natural Language Processing (EMNLP-09)*, Singapore, 2009.
- Bayesian learning of a tree substitution grammar
Matt Post and Daniel Gildea. In *Proc. Association for Computational Linguistics (short paper)*, Singapore, 2009.
- Efficient multi-pass decoding for synchronous context free grammars
Hao Zhang and Daniel Gildea. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, pages 209–217, Columbus, Ohio, 2008.
- Bayesian learning of non-compositional phrases with synchronous parsing
Hao Zhang, Chris Quirk, Robert C. Moore, and Daniel Gildea. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL-08)*, pages 97–105, Columbus, Ohio, 2008.
- Extracting synchronous grammar rules from word-level alignments in linear time
Hao Zhang, Daniel Gildea, and David Chiang. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING-08)*, pages 1081–1088, Manchester, UK, 2008.
- Parsers as language models for statistical machine translation
Matt Post and Daniel Gildea. In *The Eighth Conference of the Association for Machine Translation in the Americas (AMTA 2008)*, Honolulu, HI, 2008.
- Optimizing grammars for minimum dependency length
Daniel Gildea and David Temperley. In *Proceedings of the 45th Annual Conference of the Association for Computational Linguistics (ACL-07)*, pages 184–191, Prague, 2007.
- Worst-case synchronous grammar rules
Daniel Gildea and Daniel Štefankovič. In *Proceedings of the 2007 Meeting of the North American chapter of the Association for Computational Linguistics (NAACL-07)*, pages 147–154, 2007.
- Source-language features and maximum correlation training for machine translation evaluation

Ding Liu and Daniel Gildea. In *Proceedings of the 2007 Meeting of the North American chapter of the Association for Computational Linguistics (NAACL-07)*, pages 41–48, Rochester, NY, 2007.

Efficient search for inversion transduction grammar

Hao Zhang and Daniel Gildea. In *2006 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 224–231, Sydney, 2006.

Inducing word alignments with bilexical synchronous trees

Hao Zhang and Daniel Gildea. In *Proceedings of the International Conference on Computational Linguistics/Association for Computational Linguistics (COLING/ACL-06) Poster Session*, pages 953–960, Sydney, 2006.

Factoring synchronous grammars by sorting

Daniel Gildea, Giorgio Satta, and Hao Zhang. In *Proceedings of the International Conference on Computational Linguistics/Association for Computational Linguistics (COLING/ACL-06) Poster Session*, pages 279–286, Sydney, 2006.

Stochastic iterative alignment for machine translation evaluation

Ding Liu and Daniel Gildea. In *Proceedings of the International Conference on Computational Linguistics/Association for Computational Linguistics (COLING/ACL-06) Poster Session*, pages 539–546, Sydney, 2006.

Synchronous binarization for machine translation

Hao Zhang, Liang Huang, Daniel Gildea, and Kevin Knight. In *Proceedings of the 2006 Meeting of the North American chapter of the Association for Computational Linguistics (NAACL-06)*, pages 256–263, New York, NY, 2006.

Stochastic lexicalized inversion transduction grammar for alignment

Hao Zhang and Daniel Gildea. In *Proceedings of the 43rd Annual Conference of the Association for Computational Linguistics (ACL-05)*, pages 475–482, Ann Arbor, MI, 2005.

Dependencies vs. constituents for tree-based alignment

Daniel Gildea. In *2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 214–221, Barcelona, 2004.

Syntax-based alignment: Supervised or unsupervised?

Hao Zhang and Daniel Gildea. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING-04)*, pages 418–424, Geneva, Switzerland, August 2004.

Skeletons in the parser: Using a shallow parser to improve deep parsing

Mary Swift, James Allen, and Daniel Gildea. In *Proceedings of the 20th International Conference on Computational Linguistics (COLING-04)*, pages 383–389, Geneva, Switzerland, August 2004.

A smorgasbord of features for statistical machine translation

Franz Josef Och, Daniel Gildea, Sanjeev Khudanpur, Anoop Sarkar, Kenji Yamada, Alex Fraser, Shankar Kumar, Libin Shen, David Smith, Katherine Eng, Viren Jain, Zhen Jin, and Dragomir Radev. In *Proceedings of the 2004 Meeting of the North American chapter of the Association for Computational Linguistics (NAACL-04)*, pages 161–168, Boston, 2004.

Loosely tree-based alignment for machine translation

Daniel Gildea. In *Proceedings of the 41th Annual Conference of the Association for Computational Linguistics (ACL-03)*, pages 80–87, Sapporo, Japan, 2003.

Identifying semantic roles using Combinatory Categorical Grammar

Daniel Gildea and Julia Hockenmaier. In *2003 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 57–64, Sapporo, Japan, 2003.

- An algorithm for word-level alignment of parallel dependency trees
Yuan Ding, Daniel Gildea, and Martha Palmer. In *The 9th Machine Translation Summit of the International Association for Machine Translation*, pages 95–101, New Orleans, 2003.
- The necessity of syntactic parsing for predicate argument recognition
Daniel Gildea and Martha Palmer. In *Proceedings of the 40th Annual Conference of the Association for Computational Linguistics (ACL-02)*, pages 239–246, Philadelphia, PA, 2002.
- Probabilistic models of verb-argument structure
Daniel Gildea. In *Proceedings of the 19th International Conference on Computational Linguistics (COLING-02)*, pages 308–314, Taipei, 2002.
- Corpus variation and parser performance
Daniel Gildea. In *2001 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 167–202, Pittsburgh, PA, 2001.
- Automatic labeling of semantic roles
Daniel Gildea and Daniel Jurafsky. In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics (ACL-00)*, pages 512–520, Hong Kong, October 2000.
- Topic-based language models using EM
Daniel Gildea and Thomas Hofmann. In *Proceedings of the 6th European Conference on Speech Communication and Technology (EUROSPEECH-99)*, pages 2167–2170, Budapest, 1999.
- Forms of English function words – Effects of disfluencies, turn position, age and sex, and predictability
Alan Bell, Daniel Jurafsky, Eric Fosler-Lussier, Cynthia Girand, and Daniel Gildea. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS-99)*, pages 395–398, San Francisco, California, 1999.
- A dynamic model of aspectual composition
Nancy Chang, Daniel Gildea, and Srin Narayanan. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society (COGSCI-98)*, pages 226–231, Madison, WI, 1998.
- Applications of technology in providing transit information
Daniel Gildea and Mikael Sheikh. In *Transportation Research Record 1521*, pages 71–76, Washington, DC, 1996.
- Automatic induction of finite state transducers for simple phonological rules
Daniel Gildea and Daniel Jurafsky. In *Proceedings of the 33rd Annual Conference of the Association for Computational Linguistics (ACL-95)*, pages 9–15, Cambridge, MA, 1995.

Refereed Workshop Papers

- SCFG latent annotation for machine translation
Tagyoung Chung, Licheng Fang, and Daniel Gildea. In *Proc. of the International Workshop on Spoken Language Translation (IWSLT)*, 2011.
- Factors affecting the accuracy of Korean parsing
Tagyoung Chung, Matt Post, and Daniel Gildea. In *NAACL Workshop on Statistical Parsing of Morphologically Rich Languages*, 2010.
- Language modeling with tree substitution grammars
Matt Post and Daniel Gildea. In *NIPS workshop on Grammar Induction, Representation of Language, and Language Learning*, Whistler, British Columbia, 2009.
- Improved tree-to-string transducers for machine translation

Ding Liu and Daniel Gildea. In *ACL Workshop on Statistical Machine Translation (ACL08-SMT)*, pages 62–69, Columbus, Ohio, 2008.

Factorization of synchronous context-free grammars in linear time

Hao Zhang and Daniel Gildea. In *NAACL Workshop on Syntax and Structure in Statistical Translation (SSST)*, pages 25–32, 2007.

Syntactic features for evaluation of machine translation

Ding Liu and Daniel Gildea. In *ACL 2005 Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 25–32, Ann Arbor, MI, 2005.

Machine translation as lexicalized parsing with hooks

Liang Huang, Hao Zhang, and Daniel Gildea. In *International Workshop on Parsing Technologies (IWPT05)*, pages 65–73, 2005.

Online statistics for a unification-based dialogue parser

Micha Elsner, Mary Swift, James Allen, and Daniel Gildea. In *International Workshop on Parsing Technologies (IWPT05)*, pages 198–199, Vancouver, BC, 2005.

Invited Papers

Identifying semantic roles in text

Daniel Gildea and Daniel Jurafsky. In *Seventeenth International Joint Conference on Artificial Intelligence (IJCAI-01)*, Seattle, Washington, 2001. Appears as chapter in *Exploring Artificial Intelligence in the New Millennium*, Gerhard Lakemeyer and Bernhard Nebel, eds., Morgan Kaufmann, 2003.

Technical Reports

Improving the performance of giza++ using variational bayes

Darcey Riley and Daniel Gildea. Technical Report 963, University of Rochester, 2010.

Comparing optimization criteria and methods for word alignment

Tagyoung Chung, Shaojun Zhao, and Daniel Gildea. Technical Report 958, University of Rochester, 2010.

Bayesian learning of tokenization for machine translation

Tagyoung Chung and Daniel Gildea. Technical Report 957, University of Rochester, 2010.

Topic models for corpus-centric knowledge generalization

Benjamin Van Durme and Daniel Gildea. Technical Report 946, University of Rochester, 2009.

Syntax for statistical machine translation

Franz Josef Och, Daniel Gildea, Sanjeev Khudanpur, Anoop Sarkar, Kenji Yamada, Alex Fraser, Shankar Kumar, Libin Shen, David Smith, Katherine Eng, Viren Jain, Zhen Jin, and Dragomir Radev. Technical report, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 2003. Summer Workshop Final Report.

Natural language generation in the context of machine translation

Jan Hajič, Martin Čmejrek, Bonnie Dorr, Yuan Ding, Jason Eisner, Daniel Gildea, Terry Koo, Kristen Parton, Gerald Penn, Dragomir Radev, and Owen Rambow. Technical report, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 2002. Summer Workshop Final Report.

Topic-based novelty detection

James Allan, Hubert Jin, Martin Rajman, Charles Wayne, Daniel Gildea, Victor Lavrenko, Rose Hober-

man, and David Caputo. Technical report, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, 1999. Summer Workshop Final Report.

TEACHING EXPERIENCE

- Fall 2008 Graduate seminar on Optimization, U. of Rochester.
- Spring 2005, 2006, 2007, 2008 Graduate Machine Learning, U. of Rochester.
- Fall 2006 Graduate seminar on Language, Parsing, and Complexity, U. of Rochester.
- Fall 2005, 2007 Graduate Statistical Speech and Language Processing, U. of Rochester.
- Fall 2004 Undergraduate Design and Analysis of Efficient Algorithms, U. of Rochester.
- Spring 2004 Graduate seminar on Machine Translation, U. of Rochester.
- Fall 2002 Cotaught graduate seminar on Information Extraction, U. of Pennsylvania.

STUDENTS ADVISED

Graduated PhD students as primary adviser: Shan He (now at Google), Hao Zhang (now at Google), Ding Liu (now at Nuance).

Current PhD students at Rochester as primary ad-visor: Matt Post, Shaojun Zhao, Tagyoung Chung, Licheng Fang.

Past PhD Thesis Committees at Rochester: Chengliang Zhang (graduated 2007), Constantin Rothkopf (2008), Fabrizio Morbini, Benjamin Van Durme (2009)

Current PhD Thesis Committees at Rochester: Carlos Gomez, Bin Wei, Daphne Liu, Qi Ge, Satyaki Mahalanabis, Mehdi Manshadi, Yi Chu, Chetan Bhole, Naushad UzZaman.

Outside PhD Thesis Committees: Marta Alegre (Technical U. of Catalonia, graduated August 2004), Edward Loper (U. of Pennsylvania, graduated September 2008).

Undergraduate research advising at Rochester: Micha Elsner.

SPONSORED RESEARCH

Semantics for Statistical Machine Translation, NSF Faculty Early Career Development (CAREER) Award funded 2005-2010.

Probabilistic Tree Transducers for Machine Translation and Natural Language Processing, NSF ITR grant funded 2004-2009, Kevin Knight (USC/ISI, PI), Daniel Marcu (USC/ISI, co-PI), Daniel Gildea (U. of Rochester, co-PI).

Domain-Independent Semantic Interpretation, NSF ITR grant funded 2003-2007, Daniel Jurafsky (U. of Colorado, PI), Daniel Gildea (U. of Rochester, co-PI), Wayne Ward (U. of Colorado, co-PI), Martha Palmer (U. of Pennsylvania, co-PI), Charles Fillmore (UC Berkeley, co-PI).

PROFESSIONAL ACTIVITIES

- Editorial Board, *Computational Linguistics* (2008-2010)
- Reviewer, *Speech Communication*, *ACM Transactions on Speech and Language Processing*, *Journal of Artificial Intelligence Research*.
- Area Chair, Annual Meeting of the Association for Computational Linguistics (ACL 2008)
- Reviewer, Annual Meeting of the Association for Computational Linguistics (ACL 2000, 2003, 2005, 2007, 2008, 2009, 2010)
- Program Committee, Conference on Empirical Methods in Natural Language Processing (EMNLP 2002, 2005, 2006, 2007, 2008, 2009)
- Reviewer, Conference of the Association for Machine Translation in the Americas (AMTA 2008)
- Local Arrangements Co-chair, Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2007)
- Reviewer, American Assoc. for Artificial Intelligence conference (AAAI 2006, 2008)
- Reviewer, Conference on Natural Language Learning (CoNLL 2008)
- Reviewer, Advances in Neural Information Processing (NIPS 2004, 2005, 2006, 2007)
- Reviewer, International Joint Conference on Natural Language Processing (IJCNLP 2004, 2007)
- Reviewer, Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2004, 2006)
- Program Committee, International Workshop on Scalable Natural Language Understanding (ScaNaLU 2004, 2006)
- Co-organizer, Conference on Natural Language Learning (CoNLL 2005)
- Panelist, Doctoral Consortium, American Assoc. for Artificial Intelligence (AAAI 2005)
- Reviewer, International Conference on Computational Linguistics (COLING 2004, 2010)
- Program Committee, *Journal of Natural Language Engineering* special issue on Parallel Texts (2004)
- Reviewer, *Machine Learning Journal*, Special Issue on Learning in Speech and Language Technologies (2004)
- Reviewer, *Journal of Artificial Intelligence Research* (2004)
- Area Chair, Conference on Empirical Methods in Natural Language Processing (EMNLP 2004)
- Program Committee, Seventh International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+ 2004)
- Reviewer, Third International Conference on Language Resources and Evaluation (LREC 2002)
- Reviewer, *Journal of Machine Learning Research* special issue on Machine Learning Approaches to Shallow Parsing (2001)
- Reviewer, *Applied Intelligence* special issue on Neural Networks and Machine Learning for Natural Language Processing (2000)
- Reviewer, International Colloquium on Grammatical Inference (ICGI 1998, 2000)

Member, Association for Computing Machinery, Association for Computational Linguistics.

DEPARTMENT AND UNIVERSITY SERVICE

CS graduate admissions committee (2003-2006)

CS lab committee (2006-2008)

CS faculty recruiting committee (2007, 2008)

University faculty council (2003-2004, 2007-2008)

Co-organizer of the interdepartmental cognitive science colloquium (2006-2008)

HONORS AND AWARDS

NSF Faculty Early Career Development (CAREER) Award, 2005-2010

IRCS postdoctoral fellowship, U. of Pennsylvania, 2001-2003

National Defense Science and Engineering Graduate Fellowship, 1997-2000

National Science Foundation Graduate Fellowship (declined), 1997

California Microelectronics Graduate Fellowship, 1996-1997