

Curriculum Vitae

Brian E. Roark

Work Address:

Center for Spoken Language Understanding
OGI School of Science & Engineering
Oregon Health & Science University
20000 NW Walker Road
Beaverton, OR 97006

Contact info:

Phone: (503) 748-1752
Fax: (503) 748-1306
<http://www.cslu.ogi.edu/people/roark>
roark@cslu.ogi.edu

Education

- Ph.D. Brown University, Cognitive and Linguistic Sciences; 2001.
Thesis: [Robust Probabilistic Predictive Syntactic Processing: Motivations, Models, and Applications](#).
Committee: Mark Johnson (supervisor), Eugene Charniak, Julie Sedivy, Frederick Jelinek.
- Sc.M. Brown University, Applied Mathematics; 2001.
- M.S. Claremont Graduate School, Information Science; 1997.
- B.A. University of California, Berkeley, Double Major in Mathematics and Philosophy; 1989.

Employment

- 07/2008 – present Associate Professor
Center for Spoken Language Understanding (CSLU),
Division of Biomedical Computer Science, Dept. of Science & Engineering,
Joint appointment in Dept. of Medical Informatics and Clinical Epidemiology,
Oregon Health & Science University
- 07/2004 – Assistant Professor
07/2008 Center for Spoken Language Understanding (CSLU) and
Dept. of Computer Science & Electrical Engineering, OGI School of Science & Engineering,
Oregon Health & Science University.
- 05/2001 – Senior Technical Staff Member
06/2004 AT&T Labs-Research, Speech Algorithms Research Department.

Research interests

Syntactic parsing of text and speech; language modeling for automatic speech recognition; supervised and unsupervised learning of language and parsing models; statistical models of human language processing; spoken language processing for diagnosis of neurodevelopmental and neurodegenerative disorders; language modeling for Augmentative and Alternative Communication (AAC)

Software

GRM library (with Cyril Allauzen and Mehryar Mohri). A general software collection for constructing and modifying weighted automata and transducers representing grammars or language models.
Binaries for research purposes can be found [here](#).

Publications

BOOKS

- B. Roark and R. Sproat. 2007. [Computational Approaches to Morphology and Syntax](#). Oxford University Press, Oxford.

REFEREED JOURNALS

- B. Roark, M. Saraclar and M.J. Collins. 2007. [Discriminative n-gram language modeling](#). *Computer Speech and Language*, 21(2), pp. 373-392.
- M. Saraclar and B. Roark. 2006. [Utterance Classification with Discriminative Language Modeling](#). *Speech Communication*, 48(3-4), pp. 276-287.
- M. Bacchiani, M. Riley, B. Roark and R. Sproat. 2006. [MAP Adaptation of Stochastic Grammars](#). *Computer Speech and Language*, 20(1), pp. 41-68.
- C. Allauzen, M. Mohri and B. Roark. 2005. [The Design Principles and Algorithms of a Weighted Grammar Library](#). *International Journal of Foundations of Computer Science*, 16(3), pp. 403-421.
- B. Roark. 2004. [Robust garden path parsing](#). *Natural Language Engineering*, 10(1), pp. 1-24.
- B. Roark. 2001. [Probabilistic top-down parsing and language modeling](#). *Computational Linguistics*, 27(2), pp. 249-276.

REFEREED BOOK CHAPTERS

- B. Roark. 2008. A survey of discriminative language modeling approaches for large vocabulary continuous speech recognition. To appear in *Large Margin and Kernel Approaches to Speech and Speaker Recognition*, Joseph Keshet and Samy Bengio, Eds. Wiley.

REFEREED CONFERENCE PUBLICATIONS

- B. Roark and K. Hollingshead. 2008. [Classifying chart cells for quadratic complexity context-free inference](#). In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING)*, pp. 745-752.
- E. Arisoy, B. Roark, I. Shafran and M. Saraclar. 2008. [Discriminative N-gram Language Modeling for Turkish](#). In *Proceedings of Interspeech 2008*, pp. 825-828.
- D. Vergyri, I. Shafran, A. Stolcke, R.R. Gadde, M. Akbacak, B. Roark and W. Wang. 2007. [The SRI/OGI 2006 Spoken Term Detection System](#). In *Proceedings of Interspeech 2007*.
- K. Hollingshead and B. Roark. 2007. [Pipeline Iteration](#). In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 952-959.
- S. Fisher and B. Roark. 2007. [The utility of parse-derived features for automatic discourse segmentation](#). In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 488-495.
- B. Roark, J.P. Hosom, M. Mitchell and J.A. Kaye. 2007. [Automatically derived spoken language markers for detecting Mild Cognitive Impairment](#). In *Proceedings of the 2nd International Conference on Technology and Aging (ICTA)*.
- M. Mohri and B. Roark. 2006. [Probabilistic Context-Free Grammar Induction Based on Structural Zeros](#). In *Proceedings of the Human Language Technology Conference of the N.American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp. 312-319.
- J. Hale, I. Shafran, L. Yung, B. Dorr, M. Harper, A. Krasnyanskaya, M. Lease, Y. Liu, B. Roark, M. Snover and R. Stewart. 2006. [PCFGs with Syntactic and Prosodic Indicators of Speech Repairs](#). In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 161-169.
- B. Roark, M. Harper, E. Charniak, B. Dorr, M. Johnson, J. Kahn, Y. Liu, M. Ostendorf, J. Hale, A. Krasnyanskaya, M. Lease, I. Shafran, M. Snover, R. Stewart and L. Yung. 2006. [SParseval: Evaluation Metrics for Parsing Speech](#). In *Proceedings of the Language Resources and Evaluation Conference (LREC)*.
- B. Roark, Y. Liu, M. Harper, R. Stewart, M. Lease, M. Snover, I. Shafran, B. Dorr, J. Hale, A. Krasnyanskaya and L. Yung. 2006. [Reranking for Sentence Boundary Detection in Conversational Speech](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*.
- K. Hollingshead, S. Fisher and B. Roark. 2005. [Comparing and Combining Finite-State and Context-Free Parsers](#). In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT-EMNLP)*, pp. 787-794.

- M.J. Collins, M. Saraclar and B. Roark. 2005. [Discriminative Syntactic Language Modeling for Speech Recognition](#). In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 507-514.
- M. Saraclar and B. Roark. 2005. [Joint Discriminative Language Modeling and Utterance Classification](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 561-564.
- C. Allauzen, M. Mohri and B. Roark. 2004. [A General Weighted Grammar Library](#). In *Proceedings of the Ninth International Conference on Implementation and Application of Automata (CIAA)*, volume 3317 of Lecture Notes in Computer Science, Springer-Verlag, pp. 23-34.
- M.J. Collins and B. Roark. 2004. [Incremental Parsing with the Perceptron Algorithm](#). In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 111-118.
- B. Roark, M. Saraclar, M.J. Collins and M. Johnson. 2004. [Discriminative language modeling with conditional random fields and the perceptron algorithm](#). In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 47-54.
- M. Bacchiani, B. Roark and M. Saraclar. 2004. [Language model adaptation with MAP estimation and the perceptron algorithm](#). In *Proceedings of the Human Language Technology Conference of the N.American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp. 21-24.
- B. Roark, M. Saraclar and M.J. Collins. 2004. [Corrective language modeling for large vocabulary ASR with the perceptron algorithm](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 749-752.
- M. Bacchiani and B. Roark. 2004. [Meta-data conditional language modeling](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 241-244.
- C. Allauzen, M. Mohri, M. Riley and B. Roark. 2004. [A Generalized Construction of Speech Recognition Transducers](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 761-764.
- S.R. Maskey, M. Bacchiani, B. Roark and R. Sproat. 2004. [Improved name recognition with meta-data dependent name networks](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 789-792.
- C. Allauzen, M. Mohri and B. Roark. 2003. [Generalized algorithms for constructing language models](#). In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 40-47.
- B. Roark and M. Bacchiani. 2003. [Supervised and unsupervised PCFG adaptation to novel domains](#). In *Proceedings of the Human Language Technology Conference of the N.American Chapter of the Association for Computational Linguistics (HLT-NAACL)*, pp. 205-212.
- M. Bacchiani and B. Roark. 2003. [Unsupervised language model adaptation](#). In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 224-227.
- B. Roark. 2002. [Markov parsing: lattice rescoring with a statistical parser](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 287-294.
- B. Roark. 2001. [Explaining vowel inventory tendencies via simulation: finding a role for quantal locations and formant normalization](#). In *Proceedings of the 31st Conference of the North East Linguistics Society (NELS 31)*, pp. 419-434.
- T. Pepinsky, K. Demuth, and B. Roark. 2001. The status of ‘filler syllables’ in children’s early speech. In *Proceedings of the 25th annual Boston University Conference on Language Development*, pp. 575-586.
- M. Johnson and B. Roark. 2000. [Compact non-left-recursive grammars using the selective left-corner transform and factoring](#). In *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*, pp. 355-361.
- B. Roark and K. Demuth. 2000. Prosodic constraints and the learner’s environment: a corpus study. In *Proceedings of the 24th annual Boston University Conference on Language Development*, pp. 597-608.

- B. Roark and M. Johnson. 1999. [Efficient probabilistic top-down and left-corner parsing](#). In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 421-428.
- B. Roark and E. Charniak. 1998. [Noun-phrase co-occurrence statistics for semi-automatic semantic lexicon construction](#). In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL)*, pp. 1110-1116.

REFEREED WORKSHOP PUBLICATIONS

- B. Roark, M. Mitchell and K. Hollingshead. 2007. [Syntactic complexity measures for detecting Mild Cognitive Impairment](#). In *Proceedings of the ACL 2007 Workshop on Biomedical Natural Language Processing (BioNLP)*, pp. 1-8.
- M. Riley, B. Roark and R. Sproat. 2003. [Good-Turing estimation from word lattices for unsupervised language model adaptation](#). In *Proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 453-458.
- B. Roark. 2002. Evaluating parser accuracy using edit distance. In *Proceedings of the LREC-2002 Workshop "Beyond PARSEVAL: Towards Improved Evaluation Measures for Parsing Systems"*, pp. 30-36.
- B. Roark. 2001. Storing automatically generated treebanks in lattices of derivations. In *Proceedings of the IRCS Workshop on Linguistic Databases*, pp. 210-218.
- B. Roark and E. Charniak. 2000. Measuring efficiency in high-accuracy, broad-coverage statistical parsing. In *Proceedings of the COLING-2000 Workshop on Efficiency in Large-scale Parsing Systems*, pp. 29-36.

INVITED BOOK REVIEWS

- B. Roark. 2007. Review of Rayner et al. "Putting Linguistics into Speech Recognition". *Computational Linguistics*, , 33(2), pp. 271-273.

OTHER PUBLICATIONS

- B. Roark. 2007. Structural Alignment for Finite-State Syntactic Processing. Technical Report CSLU-07-001, OGI School of Science & Engineering, Oregon Health & Science University.
- A.M. Cohen, J. Yang, S. Fisher, B. Roark and W.R. Hersh. 2007. The OHSU Biomedical Question Answering System Framework. To appear in *Proceedings of the Text Retrieval Conference (TREC)*.
- S. Fisher and B. Roark. 2007. [Feature expansion for query-focused supervised sentence ranking](#). In *Document Understanding (DUC-2007) Workshop Papers and Agenda*.
- A.M. Cohen, J. Yang, S. Fisher, B. Roark and W.R. Hersh. 2006. [Combining Lexicon Expansion, Information Retrieval, and Cluster-based Ranking for Biomedical Question Answering](#). In *Proceedings of the Text Retrieval Conference (TREC)*.
- S. Fisher and B. Roark. 2006. [Query-focused summarization by supervised sentence ranking and skewed word distributions](#). In *Document Understanding (DUC-2006) Workshop Papers and Agenda*.
- S. Fisher, B. Roark, J. Yang and B. Hersh. 2005. [OGI/OHSU baseline query-directed multi-document summarization system for DUC 2005](#). In *Document Understanding (DUC-2005) Workshop Papers and Agenda*, pp. 100-102.
- M. Mohri and B. Roark. 2005. Structural Zeros versus Sampling Zeros. Technical Report CSEE-05-003, OGI School of Science & Engineering, Oregon Health & Science University.

Presentations

REFEREED CONFERENCE PRESENTATIONS (UNPUBLISHED)

- B. Roark and M. Johnson. 1999. Broad-coverage predictive parsing. Paper presented at *the 12th Annual CUNY Conference on Human Sentence Processing*, New York, March.

INVITED TALKS

- Classifying chart cells for quadratic complexity context-free inference
- Signal, Speech and Language Interpretation Lab (SSLI), University of Washington, July, 2008
- The Utility of Parse-derived Features for Automatic Discourse Segmentation
- Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, Dec. 2007
- Structural alignment for finite-state syntactic processing
- Center for Language and Speech Processing (CLSP), Johns Hopkins University, Sept., 2007
- Signal, Speech and Language Interpretation Lab (SSLI), University of Washington, Oct., 2007
- Joint discriminative language modeling and utterance classification
- Microsoft Research, Redmond and CLSP, Johns Hopkins, September, 2004
- Efficient incremental beam-search parsing with generative and discriminative models
- *ACL workshop on Incremental Parsing*, Barcelona, July, 2004
- Supervised and unsupervised statistical grammar adaptation to novel domains
- OGI School of Science & Engineering, Oregon Health & Science University, October, 2003
- Incremental stochastic parsing for speech recognition
- Departments of Linguistics at: Georgetown University; University of Texas, Austin; University of Washington; University of Michigan; and UCLA. February, 2002.
- Lexical and Lexico-syntactic probabilistic language models. Brown University, IGERT Conference on Stochastic and Deterministic Approaches to Vision, Language, and Cognition, May, 2001.
- Incremental Broad-coverage Predictive Parsing
- University of Minnesota, Departments of Psychology and CSE, January, 2001.
- University of Michigan, Department of Linguistics, January, 2001.
- San Diego State University, Department of Linguistics, December, 2000.
- *The 2nd Annual Graduate Conference of the Northeastern Cognitive Science Society*, University of Pennsylvania, IRCS, May, 1999.
- Probabilistic top-down parsing and language modeling.
- AT&T Labs - Research, New Jersey, December, 2000.
- IBM Research, New York, October, 2000.
- Johns Hopkins, The Center for Language and Speech Processing, June, 2000.

Honors and Fellowships

Member, DARPA Computer Science Study Panel, 2008.

Best Paper Award, HLT-NAACL 2006, for “Probabilistic Context-Free Grammar Induction Based on Structural Zeros” (with Mehryar Mohri).

Grants

NSF BCS-0826654. DHB: Measuring Spoken Language Variability in Elderly Individuals (PI: Brian Roark). 11/01/2008 - 10/31/2011. \$750,000.

NSF IIS-0811745. RI-Small: Efficient hidden structure annotation via structural multiple-sequence alignments (PI: Brian Roark). 08/01/2008 - 07/31/2011. \$400,000.

NSF IIS-0447214. CAREER: Discriminative Syntactic Language Modeling: Automatic Feature Selection and Efficient Annotation (PI: Brian Roark). 04/01/2005 - 03/31/2010. \$500,000.

Autism Speaks. Automated Measurement of Dialogue Structure in Autism (PI: Brian Roark). 12/01/2007 - 11/30/2009. \$100,000.

Department of Defense. Computer Science Study Group Participant Proposal (PI: Brian Roark). 02/01/2008 - 01/31/2009. \$100,000.

NSF IIS-0741585. SGER: RI: Text-Based Discriminative Language Modeling (PI: Brian Roark). 09/01/2007 - 08/31/2008. \$100,000.

OHSU Medical Research Foundation. Differentiating between Autism Spectrum Disorder and Developmental Language Disorder via Story Recall (PI: Brian Roark). 12/01/2006 - 11/30/2007. \$30,000.

Intel. OHSU BAIC: Technologies for behavioral assessment and intervention (PI: Tamara Hayes). Funded as co-PI (15%), 09/25/2006 - 09/24/2007.

Oregon Center for Aging and Technology. Automated Analysis of Spoken Story Recall Tests (PI: Brian Roark). 08/01/2005 - 07/31/2006. \$50,000.

Oregon Partnership for Alzheimer's Research, Tax Checkoff Research Fund. Annotation and Automatic Approximation of Language-use Metrics for Detection of Mild Cognitive Impairment (PI: Brian Roark). 05/01/2005 - 04/30/2006. \$25,000.

NSF IIS-0313383. Objective Methods for Predicting and Optimizing Synthetic Speech Quality (PI: Jan van Santen). Funded as co-PI (2 months), 09/01/2004 - 08/31/2005.

Teaching and Research

Graduate courses taught, Dept. of Computer Science & Electrical Engineering, OGI/OHSU

- CSE662 Natural Language Processing: 2005-2007(w), 2008(f)
- CSE655 Biological and Linguistic Sequence Analysis: 2007(sp), 2008(w)
- CSE654 Text-Based Language Processing Systems: 2008(sp), 2009(w)
- CSE606 Computational Approaches to Speech and Language Disorders: 2006 (su,co-taught)

Current Ph.D. students: Nathan Bodenstab (on leave), Yongshun Chen, Aaron Dunlop, Seeger Fisher, Kristy Hollingshead

Member of completed Ph.d. Thesis committees

- Fan Yang, Oregon Health & Science University, CSEE. Defended 2008.
- Jianji Yang, Oregon Health & Science University, Biomedical Informatics. Defended 2007.
- Tom Kalt, University of Massachusetts, Amherst, Computer Science. Defended 2005.

Member, CLSP Summer Workshop group, "Parsing and Spoken Structural Event Detection."
Group leader: Mary Harper. Johns Hopkins University, summer 2005.

Professional activities and affiliations

MEMBER, IEEE Signal Processing Society Speech and Language Technical Committee (2006-2008)

EDITORIAL BOARD MEMBER, Computational Linguistics (2003-2005)

PROGRAM COMMITTEE AREA CHAIR

Annual Meeting of the Association for Computational Linguistics (ACL): 2005

Human Language Technology / N.American Chapter of the ACL (HLT-NAACL): 2004,07

ORGANIZING COMMITTEE MEMBER

Tutorials co-chair, HLT-NAACL, 2009

Student Research Workshop co-chair (Faculty advisor), ACL, 2009

Publications chair, IEEE/ACL Spoken Language Technology Workshop, 2006

Publications co-chair, HLT-NAACL, 2006

REVIEWER

Journals

- Computational Linguistics: 2001-present
- Computer Speech and Language: 2003,05,08
- IEEE Transactions on Audio, Speech and Language Processing: 2004,07,08
- IEEE Transactions on Pattern Analysis and Machine Intelligence: 2006
- International Journal of Computational Linguistics and Chinese Language Processing: 2002
- Journal of Artificial Intelligence Research: 2006
- Journal of Biomedical Informatics: 2008
- Journal of Machine Learning Research: 2005
- Machine Learning: 2003
- Natural Language Engineering: 2004
- Speech Communication: 2004,05,07

Conferences

- Annual Meeting of the
 - Association for Computational Linguistics (ACL): 2001-03, 2006-08
 - European Chapter of the ACL (EACL): 2006,09
 - N.American Chapter of the ACL (NAACL): 2003,09
- Conference on Empirical Methods in Natural Language Processing (EMNLP): 2003,05,06,08
- Human Language Technology Conference (HLT): 2003,05,09
- IEEE International Conference on Acoustics, Speech, and Signal Proc. (ICASSP): 2006-08
- International Conference on Computational Linguistics (COLING): 2004,06,08
- International Joint Conference on Natural Language Processing (IJCNLP): 2008
- Neural Information Processing Systems (NIPS): 2005,07,08

Workshops

- ACL, Software, 2005
- ACL, Incremental Parsing, 2004
- HLT-NAACL, Computationally Hard Problems and Joint Inference for NLP, 2006
- HLT-NAACL, Spoken Language Understanding for Conversational Systems, 2004
- IEEE/ACL, Spoken Language Technology (SLT), 2006,08
- IEEE, Automatic Speech Recognition and Understanding (ASRU), 2001,03,07,09
- Student Research Workshop, 2001 at NAACL; 2008 at ACL.

Granting agencies

- National Science Foundation, 2004,05,07
- Alzheimer's Association , 2008

PANELIST

HLT-NAACL Doctoral Consortium, 2007

ACL Student Research Workshop, 2007

ACADEMIC SOCIETY MEMBERSHIPS

Association for Computational Linguistics (ACL).

IEEE Signal Processing Society (SPS)