Jason Baldridge
Email: jbaldrid@mail.utexas.edu
Web: http://www.jasonbaldridge.com
Twitter: @jasonbaldridge

PROFILE

I research and teach on both theoretical and applied aspects of computational linguistics. My core research interests are formal and computational models of syntax, machine learning for natural language processing, and geotemporal grounding of natural language.

EDUCATION

Ph.D., Institute for Communicating and Collaborative Systems; University of Edinburgh **2002**
Dissertation: *Lexically Specified Derivational Control in Combinatory Categorial Grammar*. Principal advisor: M. Steedman. Secondary advisor: A. Lascarides. (Won the 2003 Beth Dissertation Prize from the European Association for Logic, Language and Information.)

M.S.E, Computer Science; M.A., Linguistics; University of Pennsylvania **1998**
Thesis: *Local Scrambling and Syntactic Asymmetries in Tagalog*. Advisor: R. Clark.

B.A. Honors, Anthropology; University of Toledo **1996**
Thesis: *Reconciling Language Diversity: The History and the Future of Language Policy in India*. Advisor: S. Metris.

EXPERIENCE

Associate Professor, Dept. of Linguistics, The University of Texas at Austin, Austin, Texas **2011-Present**
Assistant Professor, Dept. of Linguistics, The University of Texas at Austin, Austin, Texas **2005-2011**
Research, teaching and supervision in computational linguistics. Associated faculty member of the Division of Statistics and Scientific Computation.

Co-founder and Chief Scientist, People Pattern Corporation, Austin, Texas **2013-Present**
Co-founded software startup to automate market research for enterprise clients. Raised $4.75 million seed and series A funding in December 2013. 20 full time employees.

Research Associate, ICCS, University of Edinburgh; Edinburgh, United Kingdom **2002-2005**
Research and implementation for active learning and robust semantic interpretation (Stanford-Edinburgh link ROSIE project, R36763).

Director, Artifactus Ltd.; Edinburgh, United Kingdom **2001-2002**
Co-founded software startup to provide corporate content management solutions using natural language processing and automated reasoning.

Summer Research Student, NASA RIACS; Moffett Field, CA **2000,2001**
Developed system to facilitate sharing of grammar resources. Extended dialog system, including integration with eye-tracking input.

GRANTS AND AWARDS

Morris Memorial Grant, New York Community Trust **2015-2017**
PI. Title: *Spatial and Temporal Analysis of Multilingual Texts*. $151,000.

Morris Memorial Grant, New York Community Trust **2012-2014**
PI. Title: *Spatial and Temporal Analysis of Multilingual Texts*. (Co-PI: Asst. Prof. M. Lease, Co-PI: Assoc. Prof. K. Erk) $73,000.

Morris Memorial Grant, New York Community Trust **2010-2012**
PI. Title: *Spatial and Temporal Analysis of Multilingual Texts*. (Co-PI: Assoc. Prof. D. Beaver, Co-PI: Asst. Prof. K. Erk) $120,000.

Structured Modeling for Low-Density Languages; Multidisciplinary University Research Initiative, Department of Defense **2010-2015**
Co-PI. Title: *The Linguistic-Core Approach to Structured Translation and Analysis of Low-Resource Languages*. (PI:Jaime Carbonell, CMU; Co-PIs: Regina Barzilay, MIT; Kevin Knight, USC-ISI; Lori Levin, CMU; Noah Smith, CMU; Stephan Vogel, CMU) $3.75 million ($420,000 sub-contract to UT Austin).

Liberal Arts Instructional Technology Services, College of Liberal Arts, UT Austin **2010-2011**
PI. Title: *Geobrowsing for Digital Humanities*. (Co-PI: Matt Cohen, English) $34,890.

Longhorn Innovation Fund for Technology **2010-2011**
Co-PI.Title: *Enabling Data-Intensive Research and Education at UT Austin via Cloud Computing*. (PI: Matthew Lease, iSchool; Co-PI: Weijia Xu, Texas Advanced Computing Center) $94,000.

Morris Memorial Grant, New York Community Trust **2008-2010**
PI.Title: *Research on Multilingual Text Interpretation*. (Co-PI: Assoc. Prof. D. Beaver, Co-PI: Asst. Prof. K. Erk) $120,000.

Documenting Endangered Languages, BCS, National Science Foundation **2007-2009**
PI (NSF BCS 0651988). Title: *Reducing Annotation Effort in the Documentation of Languages using Machine Learning and Active Learning*. (Co-PI: Asst. Prof. K. Erk, Linguistics, UT Austin). $79,106.

Doctoral Dissertation Research, BCS, National Science Foundation **2007-2008**
PI (NSF BCS 0728656). Title: *Negative Concord in Palestinian Arabic: A Formal Comparative Study*. (Co-PI: Frederick Hoyt, UT Austin). $11,892.

Morris Memorial Grant, New York Community Trust **2006-2008**
PI.Title: *Fostering International Communications through Linguistic Studies*. (Grant obtained by previous PI, Prof. Carlota Smith.) $65,000.

Information and Intelligent Systems, National Science Foundation **2006-2008**
Co-PI (NSF IIS 0535154). Title: *Extracting and Using Discourse Structure to Resolve Anaphoric Dependencies: Combining Logico-Semantic and Statistical Approaches*. (PI: Prof. N. Asher, Philosophy, UT Austin). $249,869.

Dean's Fellowship, College of Liberal Arts, UT Austin **Fall 2007**
Award for one semester research leave.

Summer Research Assignment, Dean of Graduate Studies, UT Austin **Summer 2007**
Research topic: *Bootstrapping categorial grammar lexicons from raw text*. (Two months summer salary).

Liberal Arts Instructional Technology Grant, College of Liberal Arts, UT Austin **2006-2007**
PI. Title: *Grammar Writing GUI*. $23,460.

~FAST Tex, College of Liberal Arts, UT Austin **2006**
PI. Title: *UT Austin Natural Language Processing Suite*. $1500.

JOURNAL ARTICLES

Brown, T, J. Baldridge, M. Esteva, and W. Xu. 2012. "The substantial words are in the ground and sea: computationally linking text and geography." In *Texas Studies in Literature and Language: Linguistics and Literary Studies: Computation and Convergence*. 54(3).

Enrico, J. and J. Baldridge. 2011. "Possessor Raising, Demonstrative Raising, Quantifier Float, and Number Float in Haida." *International Journal of American Linguistics*. 77(2):185-218.

Palmer, A., T. Moon, J. Baldridge, K. Erk, E. Campbell, and T. Can. 2010. "Computational strategies for reducing annotation effort in language documentation: A case study in creating interlinear texts for Uspanteko." *Linguistic Issues in Language Technologies*. 3(4): 1-42.

Denis, P. and J. Baldridge. 2009. "Global joint models for coreference resolution and named entity classification." *Procesamiento del Lenguaje Natural*. 42.

Baldridge, J. and M. Osborne. 2008. "Active Learning and Logarithmic Opinion Pools for HPSG Parse Selection". *Natural Language Engineering.* 14(2): 191-222.

Baldridge, J., N. Asher, and J. Hunter. 2007. "Annotation for and Robust Parsing of Discourse Structure on Unrestricted Texts". *Zeitschrift für Sprachwissenschaft* 26: 213-239.

Hockenmaier, J., G. Bierner, and J. Baldridge. 2004. "Extending the coverage of a CCG System". *Research in Language and Computation* 2:165-208.

CHAPTER IN REFEREED BOOK

Baldridge, J. and F. Hoyt. To appear. "Categorial Grammar." In Artemis Alexiadou and Tibor Kiss (eds.), *Syntax - Theory and Analysis. An International Handbook*. De Gruyter.

Steedman, M. and J. Baldridge. 2011. "Combinatory Categorial Grammar." In Robert Borsley and Kersti Borjars (eds.) *Non-Transformational Syntax: Formal and Explicit Models of Grammar*. Wiley-Blackwell.

PROCEEDINGS IN REFEREED CONFERENCES

Garrette, D., C. Dyer, J. Baldridge, and N. Smith. 2015. Weakly-Supervised Grammar-Informed Bayesian CCG Parser Learning. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin, Texas.

DeLozier, G., J. Baldridge and L. London. 2015. Gazetteer-Independent Toponym Resolution Using Geographic Word Profiles. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*. Austin, Texas.

Sun, L., J. Mielens and J. Baldridge. 2014. Parsing low-resource languages using Gibbs sampling for PCFGs with latent annotations. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar.

Wing, B. and Jason Baldridge. 2014. Hierarchical Discriminative Classification for Text-Based Geolocation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar.

Garrette, D., C. Dyer, J. Baldridge, and N. Smith. 2014. Weakly-Supervised Bayesian Learning of a CCG Supertagger. In *Proceedings of the 18th Conference on Computational Natural Language Learning*. Baltimore, Maryland.

Garrette, D., J. Mielens, and J. Baldridge. 2013. Real-World Semi-Supervised Learning of POS-Taggers for Low-Resource Languages. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria.

Speriosu. M. and J. Baldridge. 2013. Text-driven toponym resolution using indirect supervision. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria.

Garrette D. and J. Baldridge. 2013. Learning a part-of-speech tagger from two hours of annotation. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics*. Atlanta, Georgia.

Scott J. and J. Baldridge. 2013. A recursive estimate for the predictive likelihood in a topic model. In *Proceedings of the 16th International Conference on Artificial Intelligence and Statistics*. Scottsdale, Arizona.

Garrette D. and J. Baldridge. 2012. Type-supervised Hidden Markov Models for Part-of-Speech Tagging with Incomplete Tag Dictionaries. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Jeju, Korea.

Roller, S., M. Speriosu, S. Rallapalli, B. Wing and J. Baldridge. 2012. Supervised Text-based Geolocation Using Language Models on an Adaptive Grid. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Jeju, Korea.

Ponvert, E., J. Baldridge and K. Erk. 2011. Simple unsupervised grammar induction from raw text with cascaded finite state models. In *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon.

Wing, B. and J. Baldridge. Simple supervised document geolocation with geodesic grids. In *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*. Portland, Oregon.

Ravi, S., J. Baldridge, and K. Knight. 2010. Minimized models and grammar-informed initialization for supertagging with highly ambiguous lexicons. In *Proceedings of 48th Annual Meeting of the Association for Computational Linguistics*. Uppsala, Sweden.

Moon, T., K. Erk, and J. Baldridge. Crouching Dirichlet, Hidden Markov Model: Unsupervised POS Tagging with Context Local Tag Generation. To appear in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Cambridge, MA.

Ponvert, P. J. Baldridge, and K. Erk. Simple Unsupervised Identification of Low-level Constituents. To appear in *Proceedings of the Fourth IEEE International Conference on Semantic Computing*. Pittsburg, PA

Baldridge, J. and A. Palmer. 2009. How well does active learning actually work? Time-based evaluation of cost-reduction strategies for language documentation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore.

Moon, T., K. Erk, and J. Baldridge. 2009. Unsupervised morphological segmentation and clustering with document boundaries. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Singapore.

Ramanujam, S. and J. Baldridge. Supertagging with factorial hidden markov models. 2009. In *Proceedings of the 23rd Pacific Asia Conference on Language, Information, and Computation*. Hong Kong.

Baldridge, J. 2008. "Weakly supervised supertagging with grammar-informed initialization." In *Proceedings of the 22nd International Conference on Computational Linguistics*, pp. 57-64 Manchester, UK.

Denis, P. and J. Baldridge. 2008. Specialized models and ranking for coreference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 660-669. Honolulu, Hawaii.

Elwell, R. and J. Baldridge. 2008. "Discourse connective argument identification with connective specific rankers." In *Proceedings of the 2nd IEEE International Conference on Semantic Computing*, pp. 198-205. Santa Clara, CA.

Hoyt, F. and J. Baldridge. 2008. "A Logical Basis for the D Combinator and Normal Form Constraints in Combinatory Categorial Grammar." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pp. 326-334. Columbus, OH.

Denis, P. and J. Baldridge. 2007. "Joint determination of anaphoricity and coreference resolution using integer programming". In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 236-243. Rochester, New York. (Won best student paper award.)

Denis, P. and J. Baldridge. 2007. "A ranking approach to pronoun resolution". In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pp. 1588-1593. Hyderabad, India.

Moon, T. and J. Baldridge. 2007. "Part-of-speech tagging for Middle English through alignment and projection of parallel diachronic texts." In *Proceedings of the Joint Meeting of the Conference on Empirical Methods in Natural Language Processing and the Conference on Computational Natural Language Learning*, pp. 390-399. Prague.

Palmer, A., E. Ponvert, J. Baldridge, and C. Smith. 2007. "A sequencing model for situation entity classification." In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pp. 896-903. Prague.

Cakıcı, R. and J. Baldridge. 2006. "Projective and non-projective Turkish parsing". I*n Proceedings of the 5th Conference on Treebanks and Linguistic Theories*, pp 19-30. Prague.

Baldridge, J. and A. Lascarides. 2005. "Probabilistic head-driven parsing for discourse structure". In *Proceedings of the 9th Conference on Computational Natural Language Learning*, pp. 96-103. Ann Arbor, Michigan.

Baldridge, J. and M. Osborne. 2004. "Active learning and the total cost of annotation". In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pp. 9-16. Barcelona, Spain.

Kruijff, G. and J. Baldridge. 2004. "Generalizing dimensionality in Combinatory Categorial Grammar". In *Proceedings of the 20th International Conference on Computational Linguistic*s, pp. 191-197. Geneva, Switzerland.

Osborne, M. and J. Baldridge. 2004. "Ensemble-based active learning for parse selection". In *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 89-96. Boston, Massachusetts.

Baldridge, J. and G. Kruijff. 2003. "Multi-Modal Combinatory Categorial Grammar". In *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 211-218. Budapest, Hungary. (Won the best paper award.)

Baldridge, J. and M. Osborne. 2003. "Active learning for HPSG parse selection". In *Proceedings of the 7th Conference on Computational Natural Language Learning*, pp 17-24. Edmonton, Canada.

Baldridge, J. and G. Kruijff. 2002. "Coupling CCG with Hybrid Logic Dependency Semantics". In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 319-326. Philadelphia, Pennsylvania.

PROCEEDINGS IN REFEREED WORKSHOPS

Schneider, N., B. O'Connor, N. Saphra, D. Bamman, M. Faruqui, N. A. Smith, C. Dyer, and J. Baldridge. 2013. A framework for (under)specifying dependency syntax without overloading annotators. In *Proceedings of the 7th Linguistic Annotation Workshop*. Sofia, Bulgaria.

Speriosu, M., N. Sudan, S. Upadhyay, and J. Baldridge. 2011. Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph. In *Proceedings of the First Workshop on Unsupervised Methods in NLP.* Edinburgh, Scotland.

Speriosu, M., T. Brown, T. Moon, J. Baldridge, and K. Erk. 2010. "Connecting Language and Geography with Region-Topic Models." In *Proceedings of the Workshop on Computational Models of Spatial Language Interpretation*. Portland, Oregon.

Palmer, A., T. Moon, and J. Baldridge. 2009. "Evaluating automation strategies in language documentation." In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, pp. 36-44. Boulder, Colorado.

Shield, A. and J. Baldridge. 2009. "A morphological analyzer for verbal aspect in American Sign Language." In *Proceedings of the 10th Meeting of the Texas Linguistics Society*. Austin, Texas.

Baldridge, J. and K. Erk. 2008. "Teaching computational linguistics to a large, diverse student body: courses, tools, and interdepartmental interaction." In *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, pp 1-9. Columbus, OH.

Bird, S., E. Klein, E. Loper and J. Baldridge. 2008. "Multidisciplinary Instruction with the Natural Language Toolkit." In *Proceedings of the Third Workshop on Issues in Teaching Computational Linguistics*, pp. 62-70. Columbus, OH.

Baldridge, J., S. Chatterjee, A. Palmer, and B. Wing. 2007. "DotCCG and VisCCG: Wiki and Programming Paradigms for Improved Grammar Engineering with OpenCCG." In *Proceedings of the Workshop on Grammar Engineering Across Frameworks 2007*, pp. 5-25. Stanford, CA.

Wing, B. and J. Baldridge. 2006. "Adaptation of data and models for probabilistic parsing of Portuguese." In *Proceedings of the 7th International Workshop on the Computational Processing of the Portuguese Language*, pp.140-149. Itatiaia, Brazil.

Baldridge, J. and A. Lascarides. 2005. "Annotating discourse structures for robust semantic interpretation." In *Proceedings of the 6th International Workshop on Computational Semantics*, pp. 17-29. Tilburg, the Netherlands.

White, M. and J. Baldridge. 2003. "Adapting chart realization to CCG." In *Proceedings of the 9th European Workshop on Natural Language Generation*. Budapest, Hungary.

J. Hockenmaier, G. Bierner, and J. Baldridge. 2000. "Providing Robustness for a CCG system." In *Proceedings of the Workshop on Linguistic Theory and Grammar Implementation (ESSLLI 2000)*, pp. 97-112. Birmingham, UK.

BOOK REVIEWS AND ENCYCLOPEDIA ARTICLES

Baldridge, J. 2010. "Categorial Grammar." In the *Cambridge Encyclopedia of the Language Sciences*. Patrick Colm Hogan (ed.).

Baldridge, J. 2004. "Book Review of Verbmobil: Foundations of Speech-to-Speech Translation, Wolfgang Wahlster (ed.)". *Natural Language Engineering* 10 (2): 200-204.

OTHER PUBLICATIONS

Baldridge, J., J. Dowding, and S. Early. 2002. "Leo: an architecture for sharing resources for unification-based grammars". In *Proceedings of the 3rd International Conference on Language Resources and Evaluation*. Las Palmas, Canary Islands.

Campana, E., J. Baldridge, J. Dowding, B. A. Hockey, R. W. Remington, and L. S. Stone. 2001. "Using eye movements to determine referents in a spoken dialogue system". In *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, pp. 1-5. ACM Digital Library. Orlando, Florida.

Baldwin, B., T. Morton, A. Bagga, J. Baldridge, R. Chandraseker, A. Dimitriadis, K. Snyder, and M. Wolska. 1998. "Description of the University of Pennsylvania CAMP System as used for coreference". In *Proceedings of the 7th Message Understanding Conference*. Baltimore, Maryland.

TECHNICAL REPORTS

Reese, B. P. Denis, N. Asher, J. Baldridge, and J. Hunter. 2007. "Reference Manual for the Analysis and Annotation of Rhetorical Structure." Department of Linguistics, the University of Texas at Austin. Available from http://comp.ling.utexas.edu/discor.

XTAG Research Group. 2001. A Lexicalized Tree Adjoining Grammar for English. Institute for Research in Cognitive Science, University of Pennsylvania. IRCS-01-03.

INVITED TALKS

You Say Data, I say Data: Decoding Data for Digital Marketers. Minnesota Interactive Marketing Association. Minneapolis, Minnesota. October 2014.

Practical Sentiment Analysis. Invited tutorial, Sentiment Analysis Symposium. New York, New York. March 2014.

Learning taggers and parsers when resources are scarce. Invited talk, Linguistics Department, Cornell University. February 2014.

Grounding text. Invited talk, Text Analytics Summit. Boston, MA. June 2013.

Scaling models for text analysis. Invited talk, Bazaarvoice BV:IO Tech Talk. Bazaarvoice, Inc., Austin, Texas. May 2013.

Explicit discourse connective argument identification with connective specific rankers, with additional thoughts. Invited talk, NSF sponsored workshop on the Penn Discourse Treebank, Institute for Research in Cognitive Science, The University of Pennsylvania. May 2012.

Text Geolocation and Dating. Invited lecture, Center for Speech and Language Processing, Johns Hopkins University. April 2012.

Human Language Technology and Where It's Headed. SXSW 2012. Joint talk given with Lillian Lee (U. Cornell). March 2012.

Using universal grammar and integer programming to improve weakly-supervised supertaggers. Invited lecture, Computational Linguistics Group, The Ohio State University. May 2010.

We don't know what we don't know: important, but often overlooked, considerations for active learning. Invited lecture, Department of Linguistics, The Ohio State University. May 2010.

Using universal grammar and integer programming to improve weakly-supervised supertaggers. Artificial Intelligence and Information Systems Seminar, University of Illinois at Urbana-Champaign, Illinois. April 2010.

We don't know what we don't know: important, but often overlooked, considerations for active learning. Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA. February 2010.

Models and methods for analyzing low-density languages. Workshop on Foundations of Machine Translation and Textual Analysis for Low Density Languages. United States Military Academy, West Point, NY. September 2009.

Connecting Language Documentation and Natural Language Processing. Linguistics Society of America, Organized session on "Computational Linguistics: Implementation of Analyses against Data." San Francisco, CA. January 2009.

Weakly Supervised to Unsupervised Acquisition of Categorial Grammars. Faculty Lunch Series, Division of Statistics and Scientific Computation, UT Austin. December 2008.

Pulling the cat out of the hat: bootstrapping for categorial grammar. Institute for Communicating and Collaborative Systems, University of Edinburgh, Edinburgh, UK. Aug 2008.

Specialized rankers and global models for coreference resolution. Department of Computational Linguistics and Phonetics, University of the Saarlands, Saarbrucken, Germany. Dec 2007.

Techniques for reducing supervision in natural language processing tasks, warts and all. DFKI Language Technology Lab, Saarbrucken, Germany. Dec 2007.

Computational Grammar Writing GUI. Liberal Arts Instructional Technology Services, UT Austin. Feb 2007.

Cutting Corpus Costs: Machine Learning and Annotation. Dept. of Informatics, Catholic University of Rio de Janeiro (PUC-RIO). Jan 2007.

Cutting Corpus Costs: Machine Learning and Annotation. Keynote talk for the 10th Conference of the Texas Linguistics Society (TLSX), UT Austin. Nov 2006.

Annotation of Discourse Structures and Discourse Parsing as Dependency Parsing. Konferenz zur Verarbeitung naturlicher Sprache (KONVENS), The University of Konstanz, Germany. October 2006. (Joint talk with N. Asher)

Data-Driven Discourse Parsing. Forum for Artificial Intelligence, Dept. of Computer Science, UT Austin. Mar 2006.

Strategies and Opportunities for Computational Discourse Structure Analysis. The 4th Workshop on Discourse Structure, UT Austin. Mar 2006.

Probabilistic Head-Driven Parsing for Discourse Structure. Dept. of Linguistics. The Ohio State University, Mar 2005. (Job talk)

Probabilistic Head-Driven Parsing for Discourse Structure. Dept. of Linguistics, UT Austin, Feb 2005. (Job talk)

Making the most of what you have: modeling and annotation in data-poor conditions. The Institute for Communicating and Collaborative Systems, The University of Edinburgh. August 2004 (Job talk)

Introduction to Multi-Modal Combinatory Categorial Grammar. The Center for Computational Linguistics, Catholic University of Rio de Janeiro (PUC-RIO). August 2004.

Getting a grip on combinators: Multi-Modal Combinatory Categorial Grammar. Constraint-based Linguistics Group, Cambridge University, October 2002.

SERVICE

Editorial board: *Computational Linguistics, Computing Science and Engineering; Linguistic Issues in Language Technology.*

Grant reviewing: *Engineering and Physical Sciences Research Council (UK); National Science Foundation: Division of Behavioral and Cognitive Sciences, Program for Documenting Endangered Languages; National Science Foundation: Program for Community-based Interoperability Networks (INTEROP).*

Journal reviewing: *Computational Linguistics; Computer Speech and Language; International Journal of American Linguistics; International Journal of Semantic Computing; Journal of English Linguistics; Language Resources and Evaluation; Linguistics; Logic, Language and Information; Natural Language Engineering; Research on Language and Computation; Traitement Automatique des Langues.*

Conference program committee/reviewing: ACL Software Workshop 2005; ACL 2007, 2008 (Area chair: syntax and parsing), 2009, 2011, 2012 (Area chair: NLP applications); COLING 2004; CoNLL Dependency Parsing 2007; CoNLL 2010, 2011; EACL 2005,2012; EMNLP 2003, 2006, 2012; GEAF 2007, 2008; HLT-NAACL 2007, 2010 (Chair:Tutorials); IEEE-ICSC 2008; IJCAI 2011; IJCNLP 2008; NLP-LING 2010; PACLING 2003; PROPOR 2006, 2010; TALN 2008; TIL 2007; TLS 2006, 2007; WECOL 2007.

Member of professional organizations: The Association for Computational Linguistics; The Linguistics Society of America.

External thesis examiner: Philip Blunsom, Department of Computer Science and Engineering, University of Melbourne, 2007; Matthew Honnibal, Graduate School of Engineering and Information Technologies, University of Sydney, 2010.

OPEN SOURCE SCIENTIFIC SOFTWARE

TextGrounder **http://code.google.com/p/textgrounder/**
Project leader. Textgrounder is a system that processes texts to identify the places and times that are mentioned in them and disambiguates them to points on Earth or on the timeline.

Toolkit for Advanced Discriminative Modeling (TADM) **http://tadm.sf.net**
Developer. High performance C++ and Python toolkit for training maximum entropy and perceptron models.

OpenCCG **http://openccg.sf.net**
Original author (now maintained by Michael White, OSU). Java parsing and realization system for Combinatory Categorial Grammar. OpenCCG has been used in at least nine research dialog systems worldwide. UT Austin

project added an improved grammar engineering environment. It has been used for teaching purposes in Ankara, Austin, Cambridge, Columbus, Edinburgh, and Saarbrucken.

OpenNLP Toolkit and Maxent Toolkit **http://opennlp.sf.net, http://maxent.sf.net**
Original author. Java tools for natural language processing and Java implementation of Generalized Iterative Scaling for training maximum entropy models. Both are widely used for research in computational linguistics.

Others: see my GitHub account: https://github.com/jasonbaldridge and the UT Austin Computational Linguistics Lab GitHub account: https://github.com/utcompling

DEPARTMENTAL COMMITTEES

2012-2013: Linguistics Department Admissions Committee.

2011-2012: Linguistics Department Executive Committee.

2010-2011: Fellowship Committee; Colloquium Committee.

2008-2010: Admissions Committee (chair); Colloquium Committee.

2007-2008: Assessment Committee; Colloquium Committee; Phonetics Search Committee.

2006-2007: Assessment Committee; Colloquium Committee; Financial Aid Committee; Syntax Search Committee; Faculty Advisor for the Texas Linguistics Society.

2005-2006: Linguistics Department Executive Committee; Colloquium Committee (chair); Syntax Search Committee; Semantics Search Committee.

ADVISING

PhD Committee Chair: Pascal Denis (w/ Nicholas Asher), 2007; Dan Garrette, expected 2015; Frederick Hoyt, 2010; Alexis Palmer (w/ Katrin Erk), 2009; Elias Ponvert, 2012; Mike Speriosu, 2013; Ben Wing, expected 2015.

PhD Committee Member: Ruifang Ge (Computer Science); Rohit Kate (Computer Science), 2007; Taesun Moon, 2012; Brian Reese, 2007; Yuk Wah Wong (Computer Science), 2007.

MA Thesis advisor: Joshua Brewster, 2010; Grant DeLozier, expected 2015. Weiwei Ding, 2011; Robert Elwell, 2008; James Evans, 2014. Nazneen Rajani (Computer Science), 2014; Srivatsan Ramanujam (Computer Science), 2009; Erik Skiles, 2012.

MA Thesis, second reader: Beverly Ann Anderson, 2008; Jinung Kim, 2006; Taesik Kim, 2007; Anna Krusanova, 2008 (French); Stephen Hilderbrand, 2009.

Research Assistants: Travis Brown (2010); Sudipta Chatterjee (2007); Grant DeLozier (2013-2015); Pascal Denis (2006-2007); Robert Elwell (2007); Dan Garrette (2013-2015); Frederick Hoyt (2006-2007); Kyle Jerro (2011-2012); Vijay John (2011-2012); Abhimanu Kumar, (2011); Jason Mielens (2012-2015); Taesun Moon (2007-2009); Alexis Palmer (2006-2009); Elias Ponvert (2006-2009); Evelyn Richter (2011); Mike Speriosu (2009-2012); Liang Sun (2012-2015); Benjamin Wing (2005-2006,2012-2015).

UT AUSTIN GRADUATE COURSES TAUGHT

Applied Natural Language Processing (LIN 386) **Spring 2013**
Course introducing programming in Scala and covering core algorithms and data structures used in natural language processing, intended for PhD students in linguistics and computer science.

Learning Grounded Models of Meaning (LIN 386) **Fall 2012**
Seminar covering recent research in models that connect language to the real world.

Applied Text Analysis (LIN 386) **Spring 2012**
Course introducing programming in Scala and covering core algorithms and data structures used in text analysis, intended for PhD students in liberal arts departments.

Computational Linguistics I (LIN 386) **Spring 2006, 2007, 2011**
Course covering core algorithms and data structures used in computational linguistics. Topics include finite-state transducers, part-of-speech tagging, chunking, parsing, categorial grammar, and computational semantics. (Co-taught with Katrin Erk in Spring 2007.)

Computational Linguistics II (LIN 386) **Fall 2005, 2006**
Course covering machine learning methods in computational linguistics and advanced natural language engineering. Topics include naive Bayes, perceptron, and maximum entropy classifiers, language models, hidden Markov models, probabilistic context-free grammars and extensions, ensemble learning.

Computational Syntax (LIN 386) **Spring 2006, Fall 2008**
Course covering computational approaches to syntactic theory and grammar engineering. Topics include the Chomsky hierarchy, Tree-Adjoining Grammars, Categorial Grammars, Head-drive Phrase Structure Grammars, compositional semantics, inheritance hierarchies, structured lexicons, and probabilistic parsing models.

Categorial Grammar (LIN 393S) **Spring 2007**
Course covering syntactic theory using categorial grammars as a basis. Topics include history and development of categorial grammars, Combinatory Categorial Grammar, Categorial Type Logics, syntax-semantics interface, syntactic extraction, coordination, gapping, scrambling, intonation, and grammar implementation.

Automated Syntax-Semantics Analysis (LIN 386) **Spring 2008**
Seminar covering automated approaches for syntactic and semantic analysis. Topics include dependency parsing, unsupervised parsing, integer linear programming, cotraining, kernel methods, semantic role labeling, vector space models and word sense disambiguation. (Co-taught with Katrin Erk.)

Semisupervised Learning for Computational Linguistics (LIN 386) **Spring 2009, Fall 2010**
Seminar covering semisupervised learning methods in the context of computational linguistics. Topics include linear models for classification, clustering, generative models including Bayesian inference, self-training and co-training, topic models, and label propagation.

Analyzing Linguistic Data (LIN 392) **Fall 2009**
Course providing a hands-on introduction to statistics for language, using the R programming language. Topics include data exploration through visualization, probability distributions, mean and standard deviation of a single dataset, comparing pairs of datasets and hypotheses:testing for statistical significance, regression modeling, and clustering and classification. (Co-taught with Katrin Erk.)

UT AUSTIN UNDERGRADUATE COURSES TAUGHT

Introduction to Computational Linguistics (LIN 350) **Fall 2006**
Course covering core algorithms and data structures used in computational linguistics. Topics include finite-state transducers, part-of-speech tagging, chunking, parsing, categorial grammar, and computational semantics.

Language and Computers (LIN 312) **Spring 2007, 2008; Fall/Spring 2009-2011, Fall 2012, Spring 2013**
Course covering the relation between language and computers. Topics include text encoding, advanced search, regular expressions, document classification, dialog systems, writer's aids, cryptography, machine translation, world knowledge, and the social context of computers with language. (Co-taught with Katrin Erk in 2007.)

Natural Language Processing (LIN 350 / CS 378) **Fall 2008; Spring 2010, 2011**
Course covering applied natural language processing models and methods. Topics include n-gram language models, part-of-speech tagging with Hidden Markov Models, classification with Maximum Entropy models, probabilistic context-free grammars, and machine translation.