

Bob Carpenter

<http://bob-carpenter.github.io/>
bcarpenter@flatironinstitute.org

Positions

2020 – now	Senior Research Scientist	Flatiron Institute
2011 – 2020	Research Scientist	Columbia University
2002 – 2011	Research Scientist / Software Engineer	Alias-i (LingPipe)
2000 – 2002	Research Scientist / Software Engineer	SpeechWorks
1996 – 2000	Research Scientist	Lucent Bell Labs
1988 – 1996	Assoc. Professor, Computational Linguistics	Carnegie Mellon University

Education

1984 – 1989	Ph.D.	Cognitive Science / Computer Science	University of Edinburgh
1981 – 1984	B.S.	Math / Computer Science	Michigan State University

Grant Funding

2019–2020	Schmidt	\$500K	PI	Schmidt Futures Innovation Fellow
2019–2022	NSF	\$630K	co-PI	RIDIR: Tool for Poststratification and Small-Area Estimation for Survey Data
2019–2020	Sloan	\$360K	co-PI	Econometric Modeling with Stan for Causality and Differential Equilibrium
2017–2020	NSF	\$980K	co-PI	CRI: Stan for the Long Run
2017–2019	ONR	\$750K	co-PI	STTR Phase II: Fast and Flexible Differential Equation Model Fitting
2016–2017	ONR	\$80K	co-PI	STTR Phase I: Fast and Flexible Differential Equation Model Fitting with Applications to Pharmacometrics
2015–2018	Sloan	\$600K	co-PI	Sustaining and Growing the Stan Community
2014–2016	NSF	\$200K	co-PI	EAGER: Advanced Machine Learning for Prediction of Preterm Birth
2014–2017	IES	\$750K	co-PI	Solving Difficult Bayesian Computation Problems in Education Research Using Stan
2012–2015	NSF	\$500K	co-PI	CRI: Stan: Scalable Software for Bayesian Modeling
2006–2008	NIH	\$750K	tech lead	Phase II SBIR: Resolving Biological Entity References
2004–2006	NIH	\$400K	tech lead	Phase I SBIR: Resolving Biological Entity References

Open Source Software

2012 – present	w. Stan Dev Team.	Stan Probabilistic Programming Language and Statistical Inference (C++, R, Python)
2003 – present	LingPipe.	Natural Language Processing API. (Java)
2002 – present		Arithmetic Coding and PPM Compression API. (Java)
2001 – 2002	w. S. Caskey, K. Dayanidhi, C. Drouin, R. Pieraccini.	Open Speech Application Framework. (Java, C)
1993 – present	w. G. Penn.	Attribute Logic Engine. Grammar Formalisms. (Prolog)

Books

in prep	<i>Probability and Statistics: A Simulation-Based Introduction.</i>
in prep	<i>Bayesian Workflow with Stan.</i> CRC/Chapman & Hall. w. A. Vehtari, R. Trangucci, A. Rajagopal, D. Simpson, M. Morris, A. Gelman, and J. Gabry.
1998	<i>Type-Logical Semantics.</i> MIT Press.
1992	<i>The Logic of Typed Feature Structures.</i> Cambridge Uni. Press.

Journal Articles

- 2018 S. Paun, B. Carpenter, J. Chamberlain, D. Hovy, U. Kruschwitz, M. Poesio. Comparing Bayesian Models of Annotation. *TACL*.
- 2017 B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, A. Riddell. Stan: A Probabilistic Programming Language. *JStatSoft* 76(1).
- 2016 R. L. Grant, D. C. Furr, B. Carpenter and A. Gelman. Fitting Bayesian item response models in Stata and Stan. *Stata Journal*.
- 2014 R. Passonneau and B. Carpenter. The Benefits of a Model of Annotation. *TACL*.
- 2008 L. Smith (w. 40 others). Overview of BioCreative II Gene Mention Recognition. *Genom. Bio*.
- 2000 w. C. H. Lee, W. Chou, J. Chu-Carroll, W. Reichl, A. Saad, and Q. Zhou. On natural language call routing. *Speech Communications*
- 1999 w. J. Chu-Carroll. Dialogue Management in Vector-Based Call Routing. *Comp. Ling*.
- 1991 The Generative Power of Categorical Grammars and Head-driven Phrase Structure Grammars with Lexical Rules. *Comp. Ling*.
- 1990 w. G. Morrill. Compositionality, Implicational Logics, and Theories of Grammar. *Ling. and Phil*.
- 1988 w. G. Gazdar, G. Pullum, E. Klein, T. Hukari, R. Levine. Category Structures. *Comp. Ling*.

Open Articles

- 2019 A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P.-C. Bürkner. Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC. *arXiv* 1903.08008.
- 2015 B. Carpenter, M. D. Hoffman, M. Brubaker, D. Lee, P. Li, and M. J. Betancourt. The Stan Math Library: Reverse-Mode Automatic Differentiation in C++. *arXiv* 1509.07164.

Book Chapters

- 2004 Orthographic Variation with Lucene. *Lucene in Action*. Manning.
- 2003 Constraint-Based Processing. *Encyclopedia of Cognitive Science*. Nature Publ.
- 2003 [Computational] Complexity. *The Oxford Handbook of Computational Linguistics*. Oxford Uni. Press.
- 2000 w. J. Chu-Carroll, R. Sproat, C. Samuelsson. Computational Linguistics. *Handbook of Linguistics*. Blackwell.
- 2000 w. C. Manning. Probabilistic Parsing Using Left-Corner Language Models. *Advances in Probabilistic and Other Parsing Technologies*. Kluwer.
- 1999 The Turing-Completeness of Multimodal Categorical Grammars. *Papers Presented to Johan van Benthem in Honor of his 50th Birthday*. ESSLLI.
- 1995 w. G. Penn. Compiling Typed Attribute-Value Logic Grammars. *Current Issues in Parsing Technologies, Vol. 2*. Kluwer.
- 1993 w. G. Penn. Sources of Disjunction in Attribute-Value Logic Constraint Resolution. *Disjunctive Reasoning in Artificial Intelligence*. Ellis Horwood.
- 1993 Skeptical and Credulous Unification with Applications to Lexical Templates and Inheritance. *Default Reasoning and Lexical Organization*. Cambridge Uni. Press.
- 1992 Lexical and Unary Rules in Categorical Grammar. *Formal Grammar: Theory and Implementation*. Oxford Uni. Press.
- 1990 w. R. Thomason. Inheritance Theory and Path-Based Reasoning. *Defeasible Reasoning and Knowledge Representation*. Kluwer.

Conference Papers

- 2014 w. S. Weber, D. Lee, F. Bois, A. Gelman, and A. Racine. Bayesian drug-disease modeling with Stan. *Abstracts of the Annual Meeting of the Population Approach Group in Europe*.
- 2012 w. M. Hoffman, A. Gelman. Stan: Scalable software for probabilistic modeling. *NIPS Probabilistic Programming Workshop*.
- 2011 Measurement, Modeling, and Sampling Error in Clinical Notes. *Clinical Notes Workshop, ICML* (invit.)

Conference Papers (cont.)

- 2009 Ground Truth, Machine Learning and the Mechanical Turk. *Turk Meetup*. (invit.)
- 2008 w. B. Baldwin, C. Cano, and L. Peshkin. Known Unknowns in Discharge Summary Mining. *AMIA i2b2 Wkshp*.
- 2008 Hierarchical Bayesian Models of Categorical Data Annotation. *NYAS ML Symp*.
- 2007 LingPipe for 99.99% Recall of Gene Mentions. *BioCreative II Wkshp*.
- 2006 Character Language Models for Chinese Word Segmentation and Named-Entity Recognition. *ACL SIGHAN*.
- 2005 w. G. Morrill. Switch Graphs for Parsing Type Logical Grammars. *IWPT*.
- 2005 Scaling High-Order Character Language Models to Gigabytes. *ACL Software*.
- 2004 Phrasal Queries with LingPipe and Lucene. *TREC*.
- 2004 w. C. Collins, G. Penn. Head-Driven Parsing for Word Lattices. *ACL*.
- 2004 LingPipe for Biological Entity Extraction. *EMBO Wkshp. on Text Mining in Mol. Bio*.
- 2002 Linguistic Methodology Meets Language Reality. *PACLing*. (invit.)
- 2002 w. S. Caskey, K. Dayanidhi, C. Drouin, R. Pieraccini. A Portable, Server-Side Dialog Framework for VoiceXML. *ICSLP*.
- 2001 w. R. Pieraccini, S. Caskey, K. Dayanidhi, M. Phillips. ETUDE: A Recursive Dialog Manager with Embedded User Interface Patterns. *ASRU*.
- 2000 w. S. Lerner, R. Pieraccini. Optimizing BNF Grammars through Source Transformations. *ICSLP*.
- 1999 Tagging and Parsing Word Graphs. *Dagstuhl*. (invit.)
- 1999 Human vs. Machine: Psycholinguistics Meets Automatic Speech Recognition. *ASRU*.
- 1999 w. G. Penn. ALE for Speech: A Machine Translation Prototype. *Eurospeech*.
- 1998 w. J. Chu-Carroll. Dialogue Management in Vector-Based Call Routing. *ACL*.
- 1998 w. J. Chu-Carroll, Wolfgang Reichl, Wu Chou. Language Modeling for Content. *ICSLP*.
- 1998 w. J. Chu-Carroll. Natural Language Call Routing: A Robust, Self-Organizing Approach. *ICSLP*.
- 1998 w. J. Chu-Carroll, W. Chou, C.-H. Lee, W. Reichl, A. Saad, and Q. Zhou. A Study on Natural Language Call Routing. *IVTTA*.
- 1998 German Clause Structure: A Type-Logical Account. *ESSLLI Wkshp. on Germanic Syntax*.
- 1997 VP Ellipsis in Type-Logical Grammar. *Roma CG Wkshp*. (invit.)
- 1997 w. C. Manning. Left-Corner Language Models and Parsing. *IWPT*.
- 1995 Clause Union and Argument Structure. *Formal Gram*. (invit.)
- 1995 The Logic of Grammar. *Amsterdam Colloq*. (invit.)
- 1995 A Type-Logical Account of Plurality. *Formal Gram*.
- 1995 w. P. King. The Complexity of Closed-World Reasoning in Constraint-Based Grammar Theories. *MOL*.
- 1995 w. Y. Qu. An Abstract Machine Architecture for Typed Attribute-Value Grammars. *IWPT*.
- 1994 Principle-Based Parsing of HPSG. *CUNY Sentence Proc*. (invit.)
- 1994 Quantification and Scoping: A Deductive Account. *WCCFL*.
- 1994 w. M. Mastroianni. Constraint-Based Morpho-Phonology. *ACL SIGPhon*.
- 1993 Interpretation of Lexical Rules in HPSG. *HPSG Wkshp*. (invit.)
- 1993 An Attribute-Value Logic for Sets. *Logic and Language*.
- 1993 Compilation of Typed Attribute-Value Logic Grammars for Parsing. *IWPT*.
- 1993 The ALE System and Grammar Engineering. *EAGLES Wkshp*. (invit.)
- 1992 Constraint-Based Reasoning, Partial vs. Total Information and the Closed-World Hypothesis. *MAIS*. (invit.)
- 1992 Plurals as Further Motivation for Lambek's Abstraction. *MOL*.
- 1991 w. C. Pollard. Inclusion, Disjointness and Choice: The Logic of Linguistic Classification. *ACL*.
- 1991 w. C. Pollard, A. Franz. Specification and Implementation of Constraint-Based Unification Grammars. *IWPT*.
- 1991 Linguistic Knowledge Representation and Reasoning with Hybrid Constraints. *AAAI Fall Symp*.

Conference Papers (cont.)

- 1991 Typed Feature Structures: an Extension of First-Order Terms. *ISLP*.
- 1990 w. C. Pollard. Extensionality in Feature Structures and Feature Logic. *Bad Teinach Conf. on Unification*.
- 1989 Typed Feature Structures: Inheritance and Unification. *Wkshp. on Inheritance and NLP*.
- 1987 Some Lexical Issues in Categorical Grammar. *NLP, Unification and Grammatical Formalisms Wkshp.*
- 1987 w. Glyn Morrill. Compositionality, Implicational Logics, and Theories of Grammar. *Logic and Ling.*
- 1986 A Theory of Types for Situation Semantics. *LAGB*.

Tutorials

- 2016 w. D. Lee and A. Gelman. Stan for Bayesian Modeling and Inference. *Memorial Sloan Kettering, Vanderbilt Uni Med School, Australian ML Summer School, New England Statistics Symposium*
- 2010 w. M. Poesio. Statistical Models of Data Coding. *LREC*.
- 1999 w. J. Chu-Carroll. Spoken Dialogue Systems. *ACL*.
- 1991 Prolog for Linguistics. *LSA Summer School*.
- 1995 Type-Logical Grammar. *ESSLLI Summer School*.
- 1991 Theory and Implementation of Attribute-Value Logics. *ESSLLI Summer School*.
- 1990 Domain Theory and Denotational Semantics. *ESSLLI Summer School*.
- 1989 Event-Based Theories of Semantics. *ESSLLI Summer School*.

Patents

- 2001 Methods and apparatus for automatic call routing including disambiguating routing decisions. Robert Lee Carpenter and Jennifer Chu-Carroll. U.S. Patent number 6269153.

University Courses

- w. Ivan Sag. Categorical Grammar and HPSG. *Dutch Grad. School in Logic and Ling. Semantics and Pragmatics. CMU Grad.*
- Natural Language Processing I, II, and III. *CMU Grad.*
- Intro. to Computational Linguistics. *CMU Undergrad.*
- w. T. Seidenfeld. Philosophy of Language. *CMU Undergrad.*
- Psycholinguistics. *CMU Undergrad.*
- Computational Theory Core. *CMU Grad.*
- Computation Theory for AI. *U. Pitt. Grad.*

Blog Posts

- <http://lingpipe-blog.com/>
- <http://andrewgelman.com/>

Students

- 2002 Kathy Baker, CMU Ph.D. (Research Scientist, NSA)
- 2000 Gerald Penn, CMU Ph.D. (Assoc. Prof., Uni. Toronto, Comp. Sci.)
- 1996 Ariel Cohen, CMU Ph.D. (Senior Lect., Ben Gurion Uni., Linguistics)
- 1988–1996 14 CMU M.S. students
- 1988–1997 2 external Ph.D. committees / examinations