

Noah A. Smith

<http://homes.cs.washington.edu/~nasmith>

PROFESSIONAL EXPERIENCE

University of Washington	2015–	Associate Professor Computer Science and Engineering Senior Data Science Fellow, eScience Institute Affiliate, Center for Statistics and the Social Sciences Adjunct, Linguistics
Carnegie Mellon University	2011–2015 2006–2011	Associate Professor (tenure awarded 7/1/2014) Assistant Professor Language Technologies Institute, School of Computer Science Affiliate, Machine Learning Department

Research internships: Microsoft Research (2004), New York University (2002), Thomson Corporation (2000), Johns Hopkins University (1999), U.S. Department of Defense (1996–7)

EDUCATION

Johns Hopkins University	2006	Ph.D. , Computer Science <i>Thesis topic:</i> unsupervised natural language parsing [177]
	2004	M.S. in Engineering , Computer Science
University of Maryland	2001	B.S. with High Honors in Computer Science
	2001	B.A. with Honors in Linguistics <i>Summa cum laude; honors theses:</i> [178–180]
University of Edinburgh	2000	Visiting student in Linguistics and Artificial Intelligence

SELECTED AWARDS

- Innovation award “to stimulate innovation among faculty from a range of disciplines and to reward some of their most terrific ideas” at UW (2016–2018)
- Finmeccanica career development chair, endowed “to acknowledge promising teaching and research potential in junior faculty members” at CMU (2011–2014)
- Paper awards: ACL 2009 best paper [94], NAACL 2015 best student paper [41], ICLP 2008 best student paper [101], SMT 2013 retrospective best paper [175], and various nominations [33, 53, 62, 84, 108, 114]
- SAS/International Institute of Forecasters award (with Bryan R. Routledge; 2010)
- Institute for Quantitative Research in Finance (“Q Group”) research award (with Shimon Kogan, Bryan R. Routledge, and Jacob S. Sagi; 2008)
- Fannie and John Hertz Foundation fellowship (2001–2006)

SELECTED PRESS

- 10 inventions predicted by *The Simpsons*. Alltime10s. May 5, 2016. [125]
- My favorite thing about the Internet? Definitely the sarcasm. *MIT Technology Review*. January 21, 2016. [125]
- Talk to me. *New Yorker*. February 9, 2015. [82]
- The machines vs. Mitt Romney: How artificial intelligence is parsing political rhetoric. *Popular Science's* Zero Moment. September 24, 2014. [59, 202]
- Addiction and seduction on Yelp: The language of food love. NPR's *All Things Considered*. April 16, 2014. [10]
- The linguist's mother lode: What Twitter reveals about slang, gender, and no-nose emoticons. *Time*. September 9, 2013. [82, 91, 154]
- Do you have a Twitter 'accent'? NPR's *Here and Now*. September 4, 2013. [154]
- Main tweet: Researchers dig into the intersection of politics and Twitter. *Time's* Swampland blog. August 14, 2013. [91, 130]
- Here's how you can use Twitter to beat the spread on sports betting. *Washington Post*, August 12, 2013. [151]
- Congress' magic words. *Washington Post's* Wonkblog. December 31, 2012. [69]
- Revealed: How China censors its social networks. *New Scientist*. March 8, 2012. [17]
- Twitterology: A new science? *New York Times*. October 30, 2011. [82]
- Noah Smith on supercomputer trivia and Twitter dialects. CBC's *Spark*. February 15, 2011. [82]
- You have an accent even on Twitter. NPR's *All Things Considered*. January 18, 2011. [82]
- Twitter a decent stand-in for public opinion polls. *Ars Technica*. May 11, 2010. [91]

RESEARCH FUNDING

*Current sponsored research (*lead PI)*

NSF	*Broad-coverage semantic parsing: linguistic representation learning from crowd-scale data (with L. Zettlemoyer)	\$1.06M	2016–20
NSF	Simulation-based hypothesis testing of socio-technical community resilience using distributed optimization and natural language processing (PI: S. Miles)	\$1.21M	2015–19
DARPA	Communicating intelligently with computers (Communicating with Computers program; subcontract to ISI)	\$1.99M	2015–19
DARPA	Analysis of rare incident-event languages (Low Resource Languages for Emergent Incidents program; subcontract to CMU)	\$161K	2015–19
DARPA	Language technology for rapid response (Low Resource Languages for Emergent Incidents program; subcontract to BBN)	\$517K	2015–19
↑ UW			
↓ CMU			
NSF	Towards effective web privacy notice and choice: a multi-disciplinary perspective (PI: N. Sadeh; 1330596)	\$3.75M	2013–17

DARPA	Structured distributed semantics: analysis and filtering of text (Deep Exploration and Filtering of Text program; PI: E. Hovy)	\$5.66M	2012–17
-------	--	---------	---------

Industry gifts, internal, and other

Bloomberg	What’s the angle? Disentangling perspective from content in the news (with A. Boydston, J. H. Gross, P. Resnik)	\$60K	2016
Facebook	Computational models of framing and perspective	\$25K	2016
Amazon	Neural networks and hyperparameter optimization for NLP (Amazon Web Services credit)	\$15K	2016
UW	Mapping the landscape of perspectives in text collections	\$400K	2016–8
Google	Toward parsing any corpus	\$76K	2015
Amazon	Social measurement from social media (Amazon Web Services credit)	\$30K	2013, 2014
Google	Utility-based language models (with B. R. Routledge)	\$76K	2013
Google	Reading is believing (with W. W. Cohen, T. M. Mitchell, and C. Faloutsos)	\$500K	2012
Amazon	Collecting and analyzing tweets for large-scale political, linguistic, and economic analysis (Amazon Web Services credit)	\$5K	2011
XSEDE	Probabilistic models for recovering latent structure in natural language from text (DBS110003 and renewals)		2011–6
Google	Economic data: hard and soft (with B. R. Routledge)	\$68K	2011
CMU	Berkman faculty development award: Text-driven forecasting of mergers (with B. R. Routledge)	\$5K	2010
HP Labs	Understanding political discourse through probabilistic models (with W. W. Cohen)	\$25K	2009
Google	Growing translation resources by modeling Wikipedia	\$75K	2008
IBM	Robust, efficient, and integrable Arabic morphosyntactic disambiguation	\$30K	2007

*Past sponsored research (*lead PI)*

NSF	*CAREER: Flexible learning for natural language processing (IIS-1054319)	\$550K	2011–16
DARPA	Machine learning for adaptable heterogeneous indexing and search (Memex program; PI: J. Schneider/A. Dubrawski)	\$3.55M	2014–15
NSF	*Data-driven, computational models for discovery and analysis of framing (Socio-Computational Systems program; co-PIs A. Boydston, J. H. Gross, P. Resnik; IIS-1211277)	\$750K	2012–15
ARO	The linguistic-core approach to structured translation and analysis of low-resource languages (PI: J. Carbonell)	\$6.25M	2010–15

IARPA	Early model-based event recognition using surrogates (Open Source Indicators program; subcontract to Virginia Tech.)	\$218K	2014–15
NSF	*An exploratory study on practical approaches for robust NLP tools with integrated annotation languages (co-PI C. Dyer; IIS-1352440)	\$100K	2013–14
NSF	*Big multilinguality for data-driven lexical semantics (BIGDATA program; co-PI C. Dyer; IIS-1251131)	\$250K	2013–14
A.P. Sloan	*Identifying corporate entities and relationships from text (co-PI B. R. Routledge)	\$124K	2013–14
IARPA	*Janus: from the past into the future (Open Source Indicators program; subcontract to Raytheon BBN)	\$233K	2012–14
U.S. Army	*Story creation and inference through Bayesian extraction (subcontract on STTR to Decisive Analytics Corporation)	\$30K	2011–12
DARPA	*Text-driven forecasting of voting behavior (N10AP20042)	\$248K	2010–11
IARPA	*Text-driven forecasting (co-PI B. R. Routledge; N10PC20222)	\$257K	2010–11
Qatar NRF	*Improved Arabic natural language processing through semisupervised and cross-lingual learning (co-PI K. Oflazer; NPRP-08-485-1-083)	\$1.05M	2009–12
NSF	*Probabilistic models for structure discovery in text (IIS-0915187)	\$450K	2009–12
NSF	An integrated cluster computing architecture for machine translation (PI: S. Vogel; IIS-0844507)	\$465K	2009–11
NSF	*Scaling up unsupervised grammar induction (IIS-0836431)	\$213K	2008–9
DARPA	*Recombination, aggregation, and visualization of information in newsworthy expressions (co-PIs A. W. Black, R. Hwa, and F. L. Crabbe; NBCH-1080004)	\$500K	2008–10
NSF	*Parsing models and algorithms for morphologically rich languages (IIS-0713265)	\$112K	2007–8
DARPA	*Computer science study panel (phase 1; HR00110110013)	\$100K	2007–8

TEACHING EXPERIENCE

Tutorials

- Natural Language Processing: Algorithms and Applications, Old and New
Three-hour invited tutorial at WSDM 2015
- [Structured Sparsity in Natural Language Processing: Models, Algorithms, and Applications](#)
Three-hour tutorial at EACL 2014 (with Mário Figueiredo, André Martins, and Dani Yogatama) and NAACL 2012 (with Mário Figueiredo and André Martins)
- [Probability and Structure in Natural Language Processing](#)
Invited course at
 - Universität Heidelberg, Germany (2014)
 - International Summer School in Language and Speech Technologies, Tarragona, Spain (2012)
 - IBM Thomas J. Watson Research Center (with Shay Cohen, 2011)
- [Sequence Models](#)
Three-hour lecture at the [Lisbon Machine Learning School](#) (2011–5)

- [Structured Prediction for Natural Language Processing](#)
Three-hour invited tutorial at ICML 2009

Graduate courses

- [Natural Language Processing](#) Winter 2016
Qualifiers course for CSE Ph.D. students
- [Algorithms for Natural Language Processing](#) Fall 2011–13
Algorithms and formalisms used in NLP and CL (with Alon Lavie and Bob Frederking)
- [Structured Prediction for Language and Other Discrete Data](#) Fall 2011, 2013
Statistical structured prediction models (co-designer & co-instructor with William Cohen, Chris Dyer)
- [Probabilistic Graphical Models](#) Fall 2010
Theory and algorithms for probabilistic graphical models (instructor); Koller & Friedman textbook
- [Language and Statistics II](#) Fall 2006–9
Statistical learning for natural language processing (designer & instructor)

Advanced undergraduate course

- [Natural Language Processing](#) Spring 2008–11, 2013–4
Broad introduction to the field (designer & instructor); Jurafsky & Martin (2nd ed.) textbook
Co-taught with Chris Dyer in 2014

Graduate seminars and labs

- [Advanced Natural Language Processing](#) Spring 2016
- [Laboratory in Natural Language Processing](#) Spring 2013–4
- [Text-Driven Forecasting](#) Fall 2009
- [Advanced Natural Language Processing Seminar](#) Spring 2009–11, Fall 2012–14

Teaching prior to Carnegie Mellon

- [Empirical Research Methods in Computer Science](#) Fall 2005
Department of CS, JHU, short course for undergraduates, co-designed and co-taught with David Smith
- [Computational Genomics: Biological Sequence Modeling](#) Fall 2004
Department of CS, JHU, short course for undergraduates, co-designed and co-taught with Roy Tromble
- [Predicting English](#) Summer 2002–3
Center for Language and Speech Processing, JHU; four-hour laboratory exercise with competitive evaluation, co-designed with Jason Eisner [174], taught by others since 2004
- [Introduction to Programming for Linguists](#) (teaching assistant) Fall 2000
Department of Linguistics, University of Maryland

ADVISING

Ph.D., completed

- [Michael Heilman](#) (2008–2011); NSF graduate fellow, PIER scholar; [75, 89, 90, 136, 167, 168, 170, 173, 192, 211, 215, 233]; research scientist at Educational Testing Service; data scientist at Civis Analytics

- [Shay B. Cohen](#) (2006–2011); ICLP 2008 best student paper; Computing Innovation postdoctoral fellowship; [16, 18, 20, 71, 81, 86, 87, 96, 101, 102, 106, 140, 160, 172, 191, 242]; postdoctoral fellow at Columbia University, then lecturer (\approx assistant professor) at the University of Edinburgh
- [Dipanjan Das](#) (2008–2012); ACL 2011 best paper; [13, 66, 67, 71, 77, 85, 88, 93, 103, 136, 138, 166, 190, 211, 229, 233, 236, 243]; research scientist at Google
- [André F. T. Martins](#)¹ (2007–2012); ICTI/Portugal scholar; ACL 2009 best paper; SCS Dissertation Award Honorable Mention; IBM Portugal *Prémio Científico*; [6, 13, 21, 66, 73, 74, 76, 80, 83, 94, 95, 103, 105, 163, 164, 171, 189, 212, 238, 243]; research scientist at Priberam
- [Kevin Gimpel](#) (2006–2012); Sandia-CMU graduate fellowship; [9, 64, 68, 72, 85, 92, 100, 102, 134, 136–138, 158, 159, 175, 188, 230, 233, 236, 239]; assistant research professor, then assistant professor at TTI-Chicago
- [Tae Yano](#)² (2007–2013); [69, 99, 130, 139, 161, 169, 187, 231, 240]; engineer at Microsoft
- [Nathan Schneider](#) (2008–2014); [11, 13, 49, 56, 63, 70, 88, 131, 133, 136, 150, 152, 153, 166, 186, 211, 222, 223, 228, 232, 233]; postdoctoral researcher at the University of Edinburgh
- [Brendan O’Connor](#) (2009–2014); [7, 17, 60, 61, 63, 75, 82, 91, 136, 150, 152, 154, 157, 165, 185, 224, 225, 233]; assistant professor at the University of Massachusetts Amherst
- [Dani Yogatama](#) (2010–2015); [12, 39, 40, 43, 54, 55, 65, 75, 121, 122, 130, 136, 149, 184, 233]; research scientist at Baidu, then Google Deepmind
- [David Bamman](#) (2011–2015); Alan J. Perlis SCS graduate student teaching award; SCS Dissertation Award Honorable Mention; [17, 35, 52, 60, 125, 126, 150, 152, 157, 183, 205, 206, 225]; assistant professor at the University of California Berkeley
- [Waleed Ammar](#)³ (2011–2016); Google fellowship in natural language understanding; [5, 48, 155, 182, 207]; research scientist at the Allen Institute for Artificial Intelligence
- [Yanchuan Sim](#) (2011–2016); A*STAR national science scholarship (Singapore); [36, 45, 47, 58, 59, 65, 156, 204]; research scientist at Institute for Infocomm Research

Ph.D., all-but-dissertation (completed thesis proposal or oral examinations)

- [Jeffrey Flanigan](#)⁴ (2014–); [31, 43, 53, 136, 150, 221, 233]
- [Lingpeng Kong](#)⁵ (2013–); [26, 32, 42, 49, 120, 203]

Ph.D.

- [Sam Thomson](#) (2012–); [43, 53, 124, 150, 221]
- [Swabha Swayamdipta](#)³ (2013–); [29, 49, 150]
- [Jesse Dodge](#) (2013–); NAACL 2015 best student paper; [41, 150]
- [Dallas Card](#) (2013–); NSERC postgraduate scholarship (Canada); [25, 123, 202]
- Elizabeth Clark (2015–); NSF graduate fellow
- Lucy Lin (2015–); NSF graduate fellow
- Kelvin Luu⁶ (2015–)

¹Co-advisors: Eric Xing (CMU), Mário Figueiredo and Pedro Aguiar (IST, Universidade Técnica de Lisboa).

²Co-advisor: William Cohen (CMU).

³Co-advisor: Chris Dyer (CMU).

⁴Co-advisors: Chris Dyer (CMU) and Jaime Carbonell (CMU).

⁵Co-advisor: Chris Dyer (CMU).

⁶Co-advisor: Zoran Popovic.

- George Mulcaire (2015–) [5, 147, 148]
- Maarten Sap⁷ (2015–)
- Hao Peng (2016–)

Post-doctoral

- [Chris Dyer](#) (2010–2); [15, 75, 78, 135, 155, 158, 160, 209]; assistant professor at CMU
- [Behrang Mohit](#) (2010–2), co-supervised by Kemal Oflazer, CMU–Qatar; [70, 131, 133, 228, 232]; research scientist at Ask.com
- [Fei Liu](#) (2013–5), co-advised by Norman Sadeh; [33, 43, 50, 122, 127]; assistant professor at the University of Central Florida
- [Yangfeng Ji](#) (2016–)
- [Chenhao Tan](#) (2016–)
- [Roy Schwartz](#) (2016–), co-supervised by Oren Etzioni (AI2)

M.S.

- [Cari \(Sisson\) Bader](#) (2007–8); engineer at Nuance
- [Daniel Mills](#) (2010–1); [136, 233] (on leave)
- [Victor Chahuneau](#)³ (2012–3) [10, 15, 57, 62, 64, 132, 226, 227]; pursuing a Ph.D. at Stanford
- [Rohan Ramanath](#)⁸ (2013–5) [33, 50, 127]

Student and post-doctoral visitors and independent study projects

2016: Adhi Kuncoro [26, 30]; 2015: Miguel Ballesteros [4, 26, 29, 34, 38, 118], Renato Negrinho; 2014: Zita Marinho [27]; 2013: Daniel Preoȃiuc-Pietro (Sheffield University), Lingpeng Kong; Minghui Qiu (Singapore Management University) [45, 58]; 2012: Swapna Gottipati (Singapore Management University) [58]; 2011: Dong Nguyen [162]; 2008: Aaron Phillips, Narges Sharif-Razavian, Sourish Chaudhuri [141], Severin Hacker, ThuyLinh Nguyen [84]; 2007: Daniel Rashid [143], Mengqiu Wang [108]; 2006: Thuy Linh Nguyen

Ph.D. thesis committees

Ongoing: Adam Anderson (Harvard) [206], Justin Betteridge (CMU), Qin Gao (CMU), Greg Hanneman (CMU), Aaron Jaech [147, 148], Thuylinh Nguyen [84] (CMU), Woodley Packard, Yulia Tsvetkov [39] (CMU) 2016: Manaal Faruqui (CMU) [39–41], Yangfeng Ji (Georgia Tech.) 2015: Dan Garrette (U. Texas) [37, 46, 51], Jonathan Clark [78, 135], Jayant Krishnamurthy, Ankur Parikh, Wang Ling [38]; 2014: Anil Nelakanti (Université Pierre et Marie Curie); 2013: Khalid El-Arini, Mahesh Joshi [138], Ramnath Balasubramanyan [91, 142], Vladimir Eidelman (U. Maryland); 2012: Aaron Phillips, Sanjika Hewavitharana; 2011: Ming-Wei Chang (UIUC), David Huggins-Daines, Andreas Zollmann [98, 104]; 2010: Andrew Carlson; 2009: Andrew Arnold; 2008: Ashish Venugopal [98, 104], Ying Joy Zhang

Undergraduate research advisees

UW: Nelson Liu, Nikko Rush

CMU: Rishav Bhowmick (CMU–Qatar) [70, 194], Desai Chen [13, 88, 160, 166, 193],⁹ Arash Enayati

⁷Co-advisor: Yejin Choi.

⁸Co-advisor: Norman Sadeh (CMU).

⁹Honorable mention, CRA Outstanding Undergraduate Research Award, 2010. Pursuing a Ph.D. at MIT.

(CMU–Qatar), Mohammad Haque, Dimitry Levin [97], Philip Massey [199, 219], Zack McCord, Tobi Owoputi [63], Benjamin Plaut, Naomi Saphra [152],¹⁰ Neel Shah [195], Shiladitya Sinha [151, 224], Tal Stramer [210], Matthew Thompson [196], Daniel Tasse [216, 244],¹¹ Patrick Xia [199, 219], Xiaote Zhu

High school research advisees

Katya Anderson, Lily Scherlis [64]

Full-time technical staff

- Zach Paine (2008); later at Google, Apple
- Philip Gianfortoni (2009–10); entered M.S. in Language Technologies at CMU in 2010
- Bill McDowell (2013–4)
- Michael Mordowanec (2014) [128, 220]

SERVICE

Professional organizations

- Secretary-Treasurer of [SIGDAT](#), the Association for Computational Linguistics Special Interest Group on Linguistic Data and Corpus-Based Approaches to Natural Language Processing (2012–15)

Journals

- Associate editor, *Journal of Artificial Intelligence Research* (2014–17)
- Editorial board, *Transactions of the Association for Computational Linguistics* (2012–)
- Editorial board, *Journal of Artificial Intelligence Research* (2011–14)
- Editorial board, *Computational Linguistics* (2009–11)
- Reviewing: *Artificial Intelligence Journal* (2009), *Journal of Machine Learning Research* (2010, 2008, 2004), *Language and Computation* (2008), *IEEE Intelligent Systems* (2008), *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2011), *Journal of Applied Mathematics and Computer Science* (2014), *Journal of Information Technology and Politics* (2007), *Proceedings of the National Academy of Sciences* (2010, 2007), *Computational Linguistics* (2005), *Language Resources and Evaluation* (formerly *Computers and the Humanities*; 2004)

Conferences

- Area chair, ICLR 2017
- Program co-chair, ACL 2016
- “Tagging, chunking, syntax, and parsing” area co-chair, NAACL 2015
- Co-organizer (with Claire Cardie, Anne Washington, and John Wilkerson), [NLP Unshared Task in Polinformatics 2014](#), a research competition culminating at the ACL 2014 workshop below [201]
- Co-organizer (with Cristian Danescu-Niculescu-Mizil, Jacob Eisenstein, and Kathy McKeown), [Workshop on Language Technologies and Computational Social Science](#) at ACL 2014
- Co-organizer (with Phil Blunsom and Chris Dyer), [Workshop on Twenty Years of Bitext](#) at EMNLP 2013

¹⁰Pursuing a Ph.D. at JHU.

¹¹Pursuing a Ph.D. at CMU.

- “Social media” area co-chair, ACL 2012
- “Machine learning” area chair, EMNLP 2010
- “Parsing and syntax” area chair, NAACL-HLT 2010
- Workshops co-chair, COLING 2010
- “Parsing and syntax” area co-chair, ACL-IJCNLP 2009
- Student travel awards chair, ICML 2008
- Publications co-chair, ACL-HLT 2008
- Reviewing:
 - Natural language processing:* ACL 2015, ACL 2014, ACL 2013, ACL 2011, ACL 2010, ACL 2007, COLING-ACL 2006; COLING 2008; CoNLL 2011, CoNLL 2009, CoNLL 2006; EACL 2017, EACL 2014, EACL 2009; EMNLP 2015, EMNLP 2013, EMNLP-CoNLL 2012, EMNLP 2011, EMNLP 2009, EMNLP 2008, EMNLP-CoNLL 2007, EMNLP 2006; HLT-NAACL 2009, HLT-NAACL 2007; IJCNLP 2008; IWPT 2011; NAACL 2013
 - Machine learning and artificial intelligence:* ICML 2015, ICML 2013, ICML 2012, ICML 2011, ICML 2008, ICML 2004; IJCAI 2005; NIPS 2010, NIPS 2009, NIPS 2008, NIPS 2007; UAI 2009
 - Other:* WWW 2014, WWW 2011

Funding agencies (reviewing)

- National Science Foundation (2016, 2015, 2012, 2011, 2010, 2009, 2008, 2007, 2006)
- Qatar National Research Foundation (2012, 2011)
- Fundação para a Ciência e a Tecnologia, Portugal (2008)

University, School, and Departments

- CSE faculty recruiting committee (2017, 2016)
- LTI open house organizer (2014)
- LTI MIIS program advisory board (2013–)
- Organized a university-wide seminar series, [Machine Learning and the Social Sciences](#) (2013–14)
- Co-organized a university-wide research workshop, [Machine Learning for the Social Sciences](#) (with George Loewenstein; 2012)
- LTI faculty retreat organizer (2012)
- LTI faculty search committee chair (2012)
- MLD faculty search committee (2011, 2012)
- Faculty liaison to LTI student body (2010–14)
- CMU undergraduate research grant/fellowship selection committee (2010–14)
- SCS Intelligence Seminar organizer (2008–10)
- LTI curriculum committee (2009–14)
- LTI graduate admissions committee (2007–14)
- LTI student research symposium judge (2008, 2007)

PRESENTATIONS

Peer-refereed, non-archival conferences

- International Conference on Computational Social Science (given by Dallas Card) [197]
- International Conference on Computational Social Science (given by Yanchuan Sim) [198]
- Society of Behavioral Medicine (given by Jason Colditz) [200]
- American Political Science Association, Washington, DC, 8/2014 (given by Amber Boydston) [202]
- Annual Meeting of the Comparative Agendas Project, Konstanz, Germany, 6/2014 (given by Amber Boydston) [202]
- American Political Science Association, Chicago, IL, 8/2013 (given by Justin Gross) [204]
- Digital Humanities, Omaha, NE, 7/2013 (given by David Bamman) [206]
- Linguistic Society of America, Pittsburgh, PA, 1/2011 (given by Jacob Eisenstein) [82]
- Association for Public Policy Analysis and Management, Washington, DC, 11/1999 [218]

Invited talks at workshops, research meetings, and other events

- Mathematical Analysis of Cultural Expressive Forms: Text Data, Los Angeles, CA, 5/2016
- [New Directions in Analyzing Text as Data Conference](#), New York University, 10/2015 (given by Dallas Card)
- [Machine Learning Thematic Trimester](#), Toulouse, France, 9/2015 [54, 55]
- Veteran Affairs State of NLP, Washington, DC, 9/2015 [54, 55]
- [Atlanta Computational Social Science Workshop](#), Atlanta, GA, 11/2014 [47, 59]
- Computational Linguistics in Political Science: What Have You Done for Me Lately?, Mannheim, Germany, 10–11/2014 [interdisciplinary research]
- [New Directions in Analyzing Text as Data Conference](#), Northwestern University, 10/2014 (given by Yanchuan Sim) [47]
- [ACL 2014 Workshop on Interactive Language Learning, Visualization, and Interfaces](#), Baltimore, MD, 6/2014 [interdisciplinary research]
- [Carnegie Mellon Center for Innovation and Entrepreneurship: Launch|CMU](#), Mountain View, CA, 5/2013 [17, 69, 75, 82, 91, 97, 130, 154]
- [National Science Foundation CISE CAREER workshop](#), Arlington, TX, 5/2013
- IARPA FUSE PI meeting, College Park, MD, 4/2013 [75]
- [AAAI Spring Symposium on Analyzing Microtext](#), Stanford, CA, 3/2013 [17, 63, 82, 91, 130, 136, 154]
- [World Economic Forum](#), Davos, Switzerland, 1/2013 [69, 82, 91, 97, 138, 154]
- [New Directions in Analyzing Text as Data Conference](#), Harvard University, 10/2012 (given by Tae Yano) [69]
- [Insight 3.0: The Web Seen by its Insiders](#) 7/2012, [panel on the technology industry]
- [Workshop on Multilingual Modeling at ACL](#), 7/2012 [71]
- [Lisbon Machine Learning School](#) evening lecture, 7/2012 [17, 69, 75]
- [Inducing Linguistic Structure Workshop at NAACL](#), 6/2012 [208]
- [Computer Assisted Reporting Conference](#), 2/2012 [17, 75, 82, 91, 138]
- [Tech@State: Real-Time Awareness](#), U.S. State Department, 2/2012 [17, 75, 82, 91, 97, 138]
- [Text to Text Generation Workshop at ACL](#), 6/2011 [89, 90, 93, 108, 171, 192]
- [New Directions in Text Analysis Conference](#), Harvard University, 5/2011 [79, 82, 162]

- [South by Southwest: Interactive](#), Austin, 3/2011 [panel with Philip Resnik: “Using text to predict the real world”]
- [Tracking, Transcribing, and Tagging Government: Building Digital Records for Computational Social Science](#), Center for Advanced Study in the Behavioral Sciences, Stanford University, 6/2010 [91, 97, 99, 138, 139]
- [NAACL-HLT Workshop on #SocialMedia: Computational Linguistics in a World of Social Media](#), 6/2010 [91, 97, 99, 138, 139]
- [NIPS 2009 Workshop on Approximate Learning of Large Scale Graphical Models](#), 12/2009 [92, 94, 95, 100]
- [Hadoop Summit](#) (given by Kevin Gimpel), 6/2009 [“Natural language learning with Hadoop”]
- [NAACL-HLT Workshop on Integer Linear Programming for NLP](#) (with André Martins), 6/2009 [94, 95]
- [ACL Workshop on Mobile Language Processing](#), 6/2008 [panel on mobile NLP]
- [ACL Workshop on Issues in Teaching Computational Linguistics](#), 6/2008 [panel on NLP/CL curriculum]
- [Tokyo Forum on Advanced NLP and Text Mining](#), University of Tokyo, 2/2008 [108, 109]
- CMU “Andrew’s Leap” computer science high school outreach program talk, 7/2007
- DARPA ISAT study group on Engineering Ensemble Effects, 4/2007 [113]
- [HLT-NAACL Doctoral Consortium](#), 6/2006 [panel on the academic/industrial job market]

Invited talks at academic, industry, and government research colloquia

- Xerox Research Centre Grenoble, 7/2016 [123, 202]
- Google DeepMind/University College London, 7/2016 [123, 202]
- Federal Trade Commission, 7/2016 [123, 202]
- Georgia Institute of Technology, 4/2016 [123, 202]
- New York University, 10/2015 [35, 36, 47]
- IBM T. J. Watson Research Center, 10/2015 [35, 36, 47]
- Microsoft Research, 8/2015 [35, 36, 47]
- University of Toronto, 4/2015 [8, 47, 59]
- University of Texas at Austin, 4/2015 [47, 59]
- Georgetown University, 2/2015 [47, 59]
- Duke University, 2/2015 [47, 59]
- Universität Heidelberg, 11/2014 [8, 47, 59]
- Max-Planck-Institut SWS, 11/2014 [8, 47, 59]
- University of Washington, 4/2014 [55, 59]
- Jump Trading, 3/2014
- Department of Defense, 7/2013 [17, 69, 75]
- eBay Research Labs, 5/2013 [17, 69, 75]
- Cornell University, 3/2013 [17, 69, 75]
- University of Pennsylvania, 12/2012 [17, 69, 75]
- Universitat Politècnica de Catalunya, 7/2012 [17, 69, 75]
- Korea Advanced Institute of Science and Technology, 7/2012 [17, 69, 75, 82]
- Microsoft Research, 5/2012 [17, 69, 75, 82]
- University of Washington, 5/2012 [66, 73, 76]
- University of Maryland, 3/2012 [66, 73, 76]

- Johns Hopkins University, 3/2012 [66, 73, 76]
- Carnegie Mellon University, Language Technologies Institute 25th Anniversary, 10/2011 [17, 75, 79, 82, 91, 97, 99, 138, 139]
- In-Q-Tel, 10/2011 [75, 91, 97, 138, 139]
- IBM T. J. Watson Research Center, 5/2011 [79, 82, 162]
- Toyota Technological Institute at Chicago, 3/2011 [82, 91, 138, 139, 169]
- BlackRock, 3/2011 [82, 91, 138, 139, 169]
- University of North Carolina at Chapel Hill, Political Science Department, 2/2011 [82, 91, 138, 139, 169]
- Carnegie Mellon University, Language Technologies Institute, 9/2010 [91, 97, 99, 138, 139]
- Twitter, 7/2010 [82, 91, 97, 99, 138, 139]
- Facebook, 6/2010 [82, 91, 97, 99, 138, 139]
- HP Labs, Social Computing Group, 6/2010 [82, 91, 97, 99, 138, 139]
- Stanford University, 6/2010 [82, 91, 97, 138, 139]
- University of California Berkeley, 6/2010, [82, 91, 97, 138, 139]
- University of Illinois at Urbana-Champaign, Department of Computer Science, 5/2010 [91, 97, 99, 138, 139]
- University of Texas at Austin, Department of Computer Science, 4/2010 [91, 97, 99, 138, 139]
- Universidade Técnica de Lisboa, Instituto Superior Técnico, 3/2010 [91, 97, 99, 138, 139]
- University of Edinburgh, Institute for Communicating and Collaborative Systems, 3/2010 [91, 97, 99, 138, 139]
- University of Sheffield, Department of Computer Science, 3/2010 [91, 97, 99, 138, 139]
- University of Maryland, Department of Computer Science, 11/2009 [94, 96, 97, 102, 103, 110, 114]
- Princeton University, Department of Computer Science, 10/2009 [96, 102, 110, 114]
- University of Massachusetts at Amherst, Department of Computer Science, 3/2009 [96, 102, 110, 114]
- Universidade Técnica de Lisboa, Instituto Superior Técnico, 7/2008 [108, 175, 177]
- IBM T. J. Watson Research Center, 2/2008 [106, 108, 109]
- Carnegie Mellon University, Language Technologies Institute, 12/2007 [106, 108, 109]
- Department of Defense, 8/2007 [research overview]
- Google, 4/2006 [job talk]
- Microsoft Research, 4/2006 [job talk]
- Stanford University, Department of Computer Science, 4/2006 [job talk]
- University of Illinois at Urbana-Champaign, Department of Computer Science, 3/2006 [job talk]
- University of Wisconsin–Madison, Department of Computer Sciences, 3/2006 [job talk]
- Carnegie Mellon University, Language Technologies Institute, 3/2006 [job talk]
- University of Maryland, Institute for Advanced Computer Study, 12/2005 [114]
- Brown University, Computer Science Department, 10/2005 [111]
- Carnegie Mellon University, Language Technologies Institute, 9/2005 [111]
- Carnegie Mellon University, Center for Automated Learning and Discovery, 9/2005 [114]
- University of Pittsburgh, Computer Science Department, 9/2005 [114]
- Microsoft Research, 8/2004 [115]
- Microsoft Research, 7/2004 [116]
- University of Maryland, Institute for Advanced Computer Study, 4/2002 [117]

PUBLICATIONS

Books and proceedings

- [1] Katrin Erk and Noah A. Smith, editors. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Berlin, Germany, August 2016.
- [2] Noah A. Smith. *Linguistic Structure Prediction*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool, May 2011.

Journal articles and book chapters

- [3] Dan Jurafsky, Victor Chahuneau, Bryan R. Routledge, and Noah A. Smith. Linguistic markers of status in food culture: Bourdieu’s *Distinction* in restaurant menus. *Cultural Analytics*, 2017.
- [4] Miguel Ballesteros, Chris Dyer, Yoav Goldberg, and Noah A. Smith. Greedy transition-based dependency parsing with stack LSTMs. *Computational Linguistics*, 43(1), March 2017.
- [5] Waleed Ammar, George Mulcaire, Miguel Ballesteros, Chris Dyer, and Noah A. Smith. Many languages, one parser. *Transactions of the Association for Computational Linguistics*, 4:431–444, July 2016.
- [6] André F. T. Martins, Mário A. T. Figueiredo, Pedro M. Q. Aguiar, Noah A. Smith, and Eric P. Xing. AD³: Alternating directions dual decomposition for MAP inference in graphical models. *Journal of Machine Learning Research*, 16:495–545, March 2015.
- [7] Jacob Eisenstein, Brendan O’Connor, Noah A. Smith, and Eric P. Xing. Diffusion of language change in social media. *PLoS ONE*, November 2014.
- [8] David Bamman and Noah A. Smith. Unsupervised discovery of biographical structure from text. *Transactions of the Association for Computational Linguistics*, 2(2014):363–376, October 2014.
- [9] Kevin Gimpel and Noah A. Smith. Phrase dependency machine translation with quasi-synchronous tree-to-tree features. *Computational Linguistics*, 40(2), June 2014.
- [10] Dan Jurafsky, Victor Chahuneau, Bryan R. Routledge, and Noah A. Smith. Narrative framing of consumer sentiment in online restaurant reviews. *First Monday*, 19(4), April 2014.
- [11] Nathan Schneider, Emily Danchik, Chris Dyer, and Noah A. Smith. Discriminative lexical semantic segmentation with gaps: Running the MWE gamut. *Transactions of the Association for Computational Linguistics*, 2:193–206, April 2014.
- [12] Dani Yogatama, Chong Wang, Bryan R. Routledge, Noah A. Smith, and Eric P. Xing. Dynamic models of streaming text. *Transactions of the Association for Computational Linguistics*, 2:181–192, April 2014.
- [13] Dipanjan Das, Desai Chen, André F. T. Martins, Nathan Schneider, and Noah A. Smith. Frame-semantic parsing. *Computational Linguistics*, 40(1):9–56, March 2014.
- [14] Noah A. Smith and André F. T. Martins. Linguistic structure prediction with the sparseptron. *ACM Crossroads*, 19(3):44–48, April 2013.
- [15] Victor Chahuneau, Noah A. Smith, and Chris Dyer. pycdec: A Python interface to cdec. *Prague Bulletin of Mathematical Linguistics*, 98:51–61, October 2012.
- [16] Shay B. Cohen and Noah A. Smith. Empirical risk minimization for probabilistic grammars: Sample complexity and hardness of learning. *Computational Linguistics*, 38(3), September 2012.
- [17] David Bamman, Brendan O’Connor, and Noah A. Smith. Censorship and content deletion in Chinese social media. *First Monday*, 17(3), March 2012.
- [18] Shay B. Cohen, Robert J. Simmons, and Noah A. Smith. Products of weighted logic programs. *Theory and Practice of Logic Programming*, 11(2–3):263–296, January 2011.
- [19] Jason Eisner and Noah A. Smith. Favor short dependencies: Parsing with soft and hard constraints

- on dependency length. In Harry Bunt, Paola Merlo, and Joakim Nivre, editors, *Trends in Parsing Technology: Dependency Parsing, Domain Adaptation, and Deep Parsing*, volume 43 of *Text, Speech, and Language Technology*, chapter 8, pages 121–150. Springer, January 2011.
- [20] Shay B. Cohen and Noah A. Smith. Covariance in unsupervised learning of probabilistic grammars. *Journal of Machine Learning Research*, 11:3017–3051, November 2010.
- [21] André F. T. Martins, Noah A. Smith, Eric P. Xing, Mário A. T. Figueiredo, and Pedro M. Q. Aguiar. Nonextensive information theoretic kernels on measures. *Journal of Machine Learning Research*, 10:935–975, April 2009.
- [22] Noah A. Smith. Review of *Computational Approaches to Morphology and Syntax* by Brian Roark and Richard Sproat. *Computational Linguistics*, 34(3):453–457, September 2008.
- [23] Noah A. Smith and Mark Johnson. Weighted and probabilistic context-free grammars are equally expressive. *Computational Linguistics*, 33(4):477–491, December 2007.
- [24] Philip Resnik and Noah A. Smith. The Web as a parallel corpus. *Computational Linguistics*, 29(3):349–380, September 2003.

Refereed conference publications (full-length)

- [25] Dallas Card, Justin H. Gross, Amber E. Boydston, and Noah A. Smith. Analyzing framing through the casts of characters in the news. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, November 2016. **EMNLP 2016**. Also presented at the 2nd International Conference on Computational Social Science, June 2016.
- [26] Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, and Noah A. Smith. Distilling an ensemble of greedy dependency parsers into one MST parser. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, November 2016. **EMNLP 2016**.
- [27] Zita Marinho, André F. T. Martins, Shay B. Cohen, and Noah A. Smith. Semi-supervised learning of sequence models with method of moments. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, November 2016. **EMNLP 2016**.
- [28] Yanchuan Sim, Bryan R. Routledge, and Noah A. Smith. Friends with motives: Using text to infer influence on SCOTUS. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, November 2016. **EMNLP 2016**. Also presented at the 2nd International Conference on Computational Social Science, June 2016.
- [29] Swabha Swayamdipta, Miguel Ballesteros, Chris Dyer, and Noah A. Smith. Greedy, joint syntactic-semantic parsing with stack LSTMs. In *Proceedings of the Conference on Computational Natural Language Learning*, Berlin, Germany, August 2016. **CoNLL 2016**.
- [30] Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith. Recurrent neural network grammars. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, San Diego, CA, June 2016. **NAACL 2016**.
- [31] Jeffrey Flanigan, Chris Dyer, Noah A. Smith, and Jaime Carbonell. Generation from Abstract Meaning Representation using tree transducers. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, San Diego, CA, June 2016. **NAACL 2016**.
- [32] Lingpeng Kong, Chris Dyer, and Noah A. Smith. Segmental recurrent neural networks. In *Proceedings of the International Conference on Learning Representations*, San Juan, PR, May 2016. **ICLR 2016**.
- [33] Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A. Smith, and Frederick Liu. Crowdsourcing annotations for websites’ privacy policies: Can it really work? In *Proceedings of the International World Wide Web Conference*, Montréal, Quebec, April 2016. **WWW 2016**. **Best paper finalist**.
- [34] Miguel Ballesteros, Chris Dyer, and Noah A. Smith. Improved transition-based parsing by modeling

- characters instead of words with LSTMs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015. **EMNLP 2015.**
- [35] David Bamman and Noah A. Smith. Open extraction of fine-grained political statements. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015. **EMNLP 2015.**
- [36] Yanchuan Sim, Bryan R. Routledge, and Noah A. Smith. A utility model of authors in the scientific community. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015. **EMNLP 2015.**
- [37] Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith. A supertag-context model for weakly-supervised ccg parser learning. In *Proceedings of the Conference on Computational Natural Language Learning*, Beijing, China, July 2015. **CoNLL 2015.**
- [38] Chris Dyer, Miguel Ballesteros, Wang Ling, Austin Matthews, and Noah A. Smith. Transition-based dependency parsing with stack long short-term memory. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Beijing, China, July 2015. **ACL 2015.**
- [39] Manaal Faruqui, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah A. Smith. Sparse binary word vector representations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Beijing, China, July 2015. **ACL 2015.**
- [40] Dani Yogatama, Manaal Faruqui, Chris Dyer, and Noah A. Smith. Learning word representations with hierarchical sparse coding. In *Proceedings of the International Conference on Machine Learning*, Lille, France, July 2015. **ICML 2015.**
- [41] Manaal Faruqui, Jesse Dodge, Sujay Kumar Jauhar, Chris Dyer, Eduard Hovy, and Noah A. Smith. Retrofitting word vectors to semantic lexicons. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Denver, CO, June 2015. **NAACL 2015. Best student paper award.**
- [42] Lingpeng Kong, Alexander M. Rush, and Noah A. Smith. Transforming dependencies into phrase structures. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Denver, CO, June 2015. **NAACL 2015.**
- [43] Fei Liu, Jeffrey Flanigan, Sam Thomson, Norman Sadeh, and Noah A. Smith. Toward abstractive summarization using semantic representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Denver, CO, June 2015. **NAACL 2015.**
- [44] Nathan Schneider and Noah A. Smith. A corpus and model integrating multiword expressions and supersenses. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Denver, CO, June 2015. **NAACL 2015.**
- [45] Minghui Qiu, Yanchuan Sim, Noah A. Smith, and Jing Jiang. Modeling user arguments, interactions, and attributes for stance prediction in online debate forums. In *Proceedings of the SIAM Conference on Data Mining*, Vancouver, BC, April/May 2015. **SDM 2015.**
- [46] Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith. Weakly-supervised grammar-informed Bayesian CCG parser learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Austin, TX, January 2015. **AAAI 2015.**
- [47] Yanchuan Sim, Bryan R. Routledge, and Noah A. Smith. The utility of text: The case of amicus briefs and the Supreme Court. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Austin, TX, January 2015. **AAAI 2015.**
- [48] Waleed Ammar, Chris Dyer, and Noah A. Smith. Conditional random field autoencoders for unsupervised structured prediction. In *Advances in Neural Information Processing Systems 27*, Montréal, Quebec, December 2014. **NIPS 2014.**
- Selected for oral presentation (top 5% of accepted papers).**
- [49] Lingpeng Kong, Nathan Schneider, Swabha Swayamdipta, Archana Bhatia, Chris Dyer, and Noah A.

- Smith. A dependency parser for tweets. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Doha, Qatar, October 2014. **EMNLP 2014.**
- [50] Fei Liu, Rohan Ramanath, Norman Sadeh, and Noah A. Smith. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proceedings of the International Conference on Computational Linguistics*, Dublin, Ireland, August 2014. **COLING 2014.**
- [51] Dan Garrette, Chris Dyer, Jason Baldridge, and Noah A. Smith. Weakly-supervised Bayesian learning of a CCG supertagger. In *Proceedings of the Conference on Computational Natural Language Learning*, Baltimore, MD, June 2014. **CoNLL 2014.**
- [52] David Bamman, Ted Underwood, and Noah A. Smith. A Bayesian mixed effects model of literary character. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, June 2014. **ACL 2014.**
- [53] Jeffrey Flanigan, Sam Thomson, Jaime Carbonell, Chris Dyer, and Noah A. Smith. A discriminative graph-based parser for the abstract meaning representation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, June 2014. **ACL 2014. Nominated for best paper award.**
- [54] Dani Yogatama and Noah A. Smith. Linguistic structured sparsity in text categorization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, June 2014. **ACL 2014.**
- [55] Dani Yogatama and Noah A. Smith. Making the most of bag of words: Sentence regularization with alternating direction method of multipliers. In *Proceedings of the International Conference on Machine Learning*, Beijing, China, June 2014. **ICML 2014.**
- [56] Nathan Schneider, Spencer Onuffer, Nora Kazour, Emily Danchik, Michael T. Mordowanec, Henrietta Conrad, and Noah A. Smith. Comprehensive annotation of multiword expressions in a social web corpus. In *Proceedings of the Language Resources and Evaluation Conference*, Reykjavik, Iceland, May 2014. **LREC 2014.**
- [57] Victor Chahuneau, Eva Schlinger, Chris Dyer, and Noah A. Smith. Translating into morphologically rich languages with synthetic phrases. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, October 2013. **EMNLP 2013.**
- [58] Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang, and Noah A. Smith. Learning topics and positions from Debatepedia. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, October 2013. **EMNLP 2013.**
- [59] Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, and Noah A. Smith. Measuring ideological proportions in political speeches. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, October 2013. **EMNLP 2013.**
- [60] David Bamman, Brendan O'Connor, and Noah A. Smith. Learning latent personas of film characters. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 2013. **ACL 2013.**
- [61] Brendan O'Connor, Brandon Stewart, and Noah A. Smith. Learning to extract international relations from political context. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 2013. **ACL 2013.**
- [62] Victor Chahuneau, Noah A. Smith, and Chris Dyer. Knowledge-rich morphological priors for Bayesian language models. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013. **NAACL 2013. Nominated for best paper award.**
- [63] Olutobi Owoputi, Brendan O'Connor, Chris Dyer, Kevin Gimpel, Nathan Schneider, and Noah A. Smith. Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*

- tics, Atlanta, GA, June 2013. **NAACL 2013.**
- [64] Victor Chahuneau, Kevin Gimpel, Bryan R. Routledge, Lily Scherlis, and Noah A. Smith. Word salad: Relating food prices and descriptions. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*, Jeju, Korea, July 2012. **EMNLP 2012.**
- [65] Dani Yogatama, Yanchuan Sim, and Noah A. Smith. A probabilistic model for canonicalizing named entity mentions. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Jeju, Korea, July 2012. **ACL 2012.**
- [66] Dipanjan Das, André F. T. Martins, and Noah A. Smith. An exact dual decomposition algorithm for shallow semantic parsing with constraints. In *Proceedings of the Joint Conference on Lexical and Computational Semantics*, Montréal, Québec, June 2012. ***SEM 2012.**
- [67] Dipanjan Das and Noah A. Smith. Graph-based lexicon expansion with sparsity-inducing penalties. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Montréal, Québec, June 2012. **NAACL 2012.**
- [68] Kevin Gimpel and Noah A. Smith. Structured ramp loss minimization for machine translation. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Montréal, Québec, June 2012. **NAACL 2012.**
- [69] Tae Yano, Noah A. Smith, and John D. Wilkerson. Textual predictors of bill survival in Congressional committees. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, pages 793–802, Montréal, Québec, June 2012. **NAACL 2012.**
- [70] Behrang Mohit, Nathan Schneider, Rishav Bhowmick, Kemal Oflazer, and Noah A. Smith. Recall-oriented learning of named entities in Arabic Wikipedia. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, Avignon, France, April 2012. **EACL 2012.**
- [71] Shay B. Cohen, Dipanjan Das, and Noah A. Smith. Unsupervised structure prediction with non-parallel multilingual guidance. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. **EMNLP 2011.**
- [72] Kevin Gimpel and Noah A. Smith. Quasi-synchronous phrase dependency grammars for machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. **EMNLP 2011.**
- [73] André F. T. Martins, Noah A. Smith, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Dual decomposition with many overlapping components. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. **EMNLP 2011.**
- [74] André F. T. Martins, Noah A. Smith, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Structured sparsity in structured prediction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. **EMNLP 2011.**
- [75] Dani Yogatama, Michael Heilman, Brendan O’Connor, Chris Dyer, Bryan R. Routledge, and Noah A. Smith. Predicting a scientific community’s response to an article. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, July 2011. **EMNLP 2011.**
- [76] André F. T. Martins, Pedro M. Q. Aguiar, Mário A. T. Figueiredo, Noah A. Smith, and Eric P. Xing. An augmented Lagrangian approach to constrained MAP inference. In *Proceedings of the International Conference on Machine Learning*, Bellevue, WA, June/July 2011. **ICML 2011.**
- [77] Dipanjan Das and Noah A. Smith. Semi-supervised frame-semantic parsing for unknown predicates. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Portland, OR, June 2011. **ACL 2011.**
- [78] Chris Dyer, Jonathan H. Clark, Alon Lavie, and Noah A. Smith. Unsupervised word alignment with arbitrary features. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Portland, OR, June 2011. **ACL 2011.**

- [79] Jacob Eisenstein, Noah A. Smith, and Eric P. Xing. Discovering sociolinguistic associations with structured sparsity. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Portland, OR, June 2011. **ACL 2011.**
- [80] André F. T. Martins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Online learning of structured predictors with multiple kernels. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, Fort Lauderdale, FL, April 2011. **AISTATS 2011.**
- [81] Shay B. Cohen and Noah A. Smith. Empirical risk minimization with approximations of probabilistic grammars. In *Advances in Neural Information Processing Systems 23*, Vancouver, BC, December 2010. **NIPS 2010.**
- [82] Jacob Eisenstein, Brendan O’Connor, Noah A. Smith, and Eric P. Xing. A latent variable model for geographic lexical variation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Cambridge, MA, October 2010. **EMNLP 2010.**
- [83] André F. T. Martins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Turbo parsers: Dependency parsing by approximate variational inference. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Cambridge, MA, October 2010. **EMNLP 2010.**
- [84] ThuyLinh Nguyen, Stephan Vogel, and Noah A. Smith. Nonparametric word segmentation for machine translation. In *Proceedings of the International Conference on Computational Linguistics*, Beijing, China, August 2010. **COLING 2010.**
Best paper finalist.
- [85] Kevin Gimpel, Dipanjan Das, and Noah A. Smith. Distributed asynchronous online learning for natural language processing. In *Proceedings of the Conference on Computational Natural Language Learning*, Uppsala, Sweden, July 2010. **CoNLL 2010.**
- [86] Shay B. Cohen and Noah A. Smith. Viterbi training for PCFGs: Hardness results and competitiveness of uniform initialization. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 1502–1511, Uppsala, Sweden, July 2010. **ACL 2010.**
- [87] Shay B. Cohen, David M. Blei, and Noah A. Smith. Variational inference for adaptor grammars. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. **NAACL 2010.**
- [88] Dipanjan Das, Nathan Schneider, Desai Chen, and Noah A. Smith. Probabilistic frame-semantic parsing. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. **NAACL 2010.**
- [89] Michael Heilman and Noah A. Smith. Good question! statistical ranking for question generation. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. **NAACL 2010.**
- [90] Michael Heilman and Noah A. Smith. Tree edit models for recognizing textual entailments, paraphrases, and answers to questions. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. **NAACL 2010.**
- [91] Brendan O’Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, pages 122–129, Washington, DC, May 2010. **ICWSM 2010.**
- [92] Kevin Gimpel and Noah A. Smith. Feature-rich translation by quasi-synchronous lattice parsing. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 219–228, Singapore, August 2009. **EMNLP 2009.**
- [93] Dipanjan Das and Noah A. Smith. Paraphrase identification as probabilistic quasi-synchronous recog-

- inition. In *Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 468–476, Singapore, August 2009. **ACL 2009.**
- [94] André F. T. Martins, Noah A. Smith, and Eric P. Xing. Concise integer linear programming formulations for dependency parsing. In *Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing*, pages 342–350, Singapore, August 2009. **ACL 2009.**
Best paper award.
- [95] André F. T. Martins, Noah A. Smith, and Eric P. Xing. Polyhedral outer approximations with application to natural language parsing. In *Proceedings of the International Conference on Machine Learning*, pages 713–720, Montréal, Québec, June 2009. **ICML 2009.**
- [96] Shay B. Cohen and Noah A. Smith. Shared logistic normal distributions for soft parameter tying in unsupervised grammar induction. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pages 74–82, Boulder, CO, May/June 2009. **NAACL 2009.**
- [97] Shimon Kogan, Dimitry Levin, Bryan R. Routledge, Jacob S. Sagi, and Noah A. Smith. Predicting risk from financial reports with regression. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pages 272–280, Boulder, CO, May/June 2009. **NAACL 2009.**
- [98] Ashish Venugopal, Andreas Zollmann, Noah A. Smith, and Stephan Vogel. Preference grammars: Softening syntactic constraints to improve statistical machine translation. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pages 236–244, Boulder, CO, May/June 2009. **NAACL 2009.**
- [99] Tae Yano, William W. Cohen, and Noah A. Smith. Predicting response to political blog posts with topic models. In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference*, pages 477–485, Boulder, CO, May/June 2009. **NAACL 2009.**
- [100] Kevin Gimpel and Noah A. Smith. Cube summing, approximate inference with non-local features, and dynamic programming without semirings. In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, pages 157–166, Athens, Greece, March/April 2009. **EACL 2009.**
- [101] Shay B. Cohen, Robert J. Simmons, and Noah A. Smith. Dynamic programming algorithms as products of weighted logic programs. In *Proceedings of the International Conference on Logic Programming*, Udine, Italy, December 2008. **ICLP 2008.**
Best student paper award.
- [102] Shay B. Cohen, Kevin Gimpel, and Noah A. Smith. Logistic normal priors for unsupervised probabilistic grammar induction. In *Advances in Neural Information Processing Systems 21*, pages 321–328, Vancouver, BC, December 2008. **NIPS 2008.**
- [103] André F. T. Martins, Dipanjan Das, Noah A. Smith, and Eric P. Xing. Stacking dependency parsers. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 157–166, Waikiki, HI, October 2008. **EMNLP 2008.**
- [104] Ashish Venugopal, Andreas Zollmann, Noah A. Smith, and Stephan Vogel. Wider pipelines: N -best alignments and parses in MT training. In *Proceedings of the Conference of the Association for Machine Translation in the Americas*, Waikiki, HI, October 2008. **AMTA 2008.**
- [105] André F. T. Martins, Mário A. T. Figueiredo, Pedro M. Q. Aguiar, Noah A. Smith, and Eric P. Xing. Nonextensive entropic kernels. In *Proceedings of the International Conference on Machine Learning*, pages 640–647, Helsinki, Finland, July 2008. **ICML 2008.**
- [106] Shay B. Cohen and Noah A. Smith. Joint morphological and syntactic disambiguation. In *Proceed-*

- ings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 208–217, Prague, Czech Republic, June 2007. **EMNLP-CoNLL 2007.**
- [107] David A. Smith and Noah A. Smith. Probabilistic models of nonprojective dependency trees. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 132–140, Prague, Czech Republic, June 2007. **EMNLP-CoNLL 2007.**
- [108] Mengqiu Wang, Noah A. Smith, and Teruko Mitamura. What is the Jeopardy model? a quasi-synchronous grammar for QA. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 22–32, Prague, Czech Republic, June 2007. **EMNLP-CoNLL 2007.**
Nominated for best paper award.
- [109] Noah A. Smith, Douglas L. Vail, and John D. Lafferty. Computationally efficient M-estimation of log-linear structure models. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 752–759, Prague, Czech Republic, June 2007. **ACL 2007.**
- [110] Noah A. Smith and Jason Eisner. Annealing structural bias in multilingual weighted grammar induction. In *Proceedings of the International Conference on Computational Linguistics and Annual Meeting of the Association for Computational Linguistics*, pages 569–576, Sydney, Australia, July 2006. **COLING-ACL 2006.**
- [111] Jason Eisner and Noah A. Smith. Parsing with soft and hard constraints on dependency length. In *Proceedings of the International Workshop on Parsing Technologies*, pages 30–41, Vancouver, BC, October 2005. **IWPT 2005.**
- [112] Noah A. Smith, David A. Smith, and Roy W. Tromble. Context-based morphological disambiguation with random fields. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 475–482, Vancouver, BC, October 2005. **EMNLP 2005.**
- [113] Jason Eisner, Eric Goldlust, and Noah A. Smith. Compiling Comp Ling: Practical weighted dynamic programming and the Dyna language. In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 281–290, Vancouver, BC, October 2005. **EMNLP 2005.**
- [114] Noah A. Smith and Jason Eisner. Contrastive estimation: Training log-linear models on unlabeled data. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 354–362, Ann Arbor, MI, June 2005. **ACL 2005.**
Nominated for best paper award.
- [115] David A. Smith and Noah A. Smith. Bilingual parsing with factored estimation: Using English to parse Korean. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 49–56, Barcelona, Spain, July 2004. **EMNLP 2004.**
- [116] Noah A. Smith and Jason Eisner. Annealing techniques for unsupervised statistical language learning. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pages 487–494, Barcelona, Spain, July 2004. **ACL 2004.**
- [117] Noah A. Smith. From words to corpora: Recognizing translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 95–102, Philadelphia, PA, July 2002. **EMNLP 2002.**

Refereed conference publications (short)

- [118] Miguel Ballesteros, Yoav Goldberg, Chris Dyer, and Noah A. Smith. Training with exploration improves a greedy stack LSTM parser. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, November 2016. **EMNLP 2016.**
- [119] Kazuya Kawakami, Chris Dyer, Bryan R. Routledge, and Noah A. Smith. Character sequence models for colorful words. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Austin, TX, November 2016. **EMNLP 2016.**
- [120] Liang Lu, Lingpeng Kong, Chris Dyer, Noah A. Smith, and Steve Renals. Segmental recurrent neural networks for end-to-end speech recognition. In *Proceedings of InterSpeech*, San Francisco, CA, September 2016. **InterSpeech 2016.**
- [121] Dani Yogatama, Lingpeng Kong, and Noah A. Smith. Bayesian optimization of text representations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015. **EMNLP 2015.**
- [122] Dani Yogatama, Fei Liu, and Noah A. Smith. Extractive summarization by maximizing semantic volume. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Lisbon, Portugal, September 2015. **EMNLP 2015.**
- [123] Dallas Card, Amber E. Boydston, Justin H. Gross, Philip Resnik, and Noah A. Smith. The Media Frames Corpus: Annotations of frames across issues. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Beijing, China, July 2015. **ACL 2015.**
- [124] Meghana Kshirsagar, Sam Thomson, Nathan Schneider, Jaime Carbonell, Noah A. Smith, and Chris Dyer. Frame-semantic role labeling with heterogeneous annotations. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Beijing, China, July 2015. **ACL 2015.**
- [125] David Bamman and Noah A. Smith. Contextualized sarcasm detection on twitter. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Oxford, UK, May 2015. **ICWSM 2015.**
- [126] David Bamman, Chris Dyer, and Noah A. Smith. Distributed representations of geographically situated language. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, June 2014. **ACL 2014.**
- [127] Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah A. Smith. Unsupervised alignment of privacy policies using hidden Markov models. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Baltimore, MD, June 2014. **ACL 2014.**
- [128] Michael T. Mordowanec, Nathan Schneider, Chris Dyer, and Noah A. Smith. Simplified dependency annotations with GFL-Web. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, Baltimore, MD, June 2014. **ACL 2014 demonstration track.**
- [129] André F. T. Martins, Miguel Almeida, and Noah A. Smith. Turning on the turbo: Fast third-order non-projective turbo parsers. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Sofia, Bulgaria, August 2013. **ACL 2013.**
- [130] Tae Yano, Dani Yogatama, and Noah A. Smith. A penny for your tweets: Campaign contributions and Capitol Hill microblogs. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Boston, MA, July 2013. **ICWSM 2013.**
- [131] Nathan Schneider, Behrang Mohit, Chris Dyer, Kemal Oflazer, and Noah A. Smith. Supersense tagging for Arabic: the MT-in-the-middle attack. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013. **NAACL 2013.**
- [132] Chris Dyer, Victor Chahuneau, and Noah A. Smith. A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Atlanta, GA, June 2013. **NAACL 2013.**

- [133] Nathan Schneider, Behrang Mohit, Kemal Oflazer, and Noah A. Smith. Coarse lexical semantic annotation with supersenses: An Arabic case study. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Jeju, Korea, July 2012. **ACL 2012.**
- [134] Kevin Gimpel and Noah A. Smith. Concavity and initialization for unsupervised dependency parsing. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Montréal, Québec, June 2012. **NAACL 2012.**
- [135] Jonathan H. Clark, Chris Dyer, Alon Lavie, and Noah A. Smith. Better hypothesis testing for statistical machine translation: Controlling for optimizer instability. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, Portland, OR, June 2011. **ACL 2011.**
- [136] Kevin Gimpel, Nathan Schneider, Brendan O’Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, and Noah A. Smith. Part-of-speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, Portland, OR, June 2011. **ACL 2011.**
- [137] Kevin Gimpel and Noah A. Smith. Softmax-margin CRFs: Training log-linear models with cost functions. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. **NAACL 2010.**
- [138] Mahesh Joshi, Dipanjan Das, Kevin Gimpel, and Noah A. Smith. Movie reviews and revenues: An experiment in text regression. In *Proceedings of the North American Chapter of the Association for Computational Linguistics Human Language Technologies Conference*, Los Angeles, CA, June 2010. **NAACL 2010.**
- [139] Tae Yano and Noah A. Smith. What’s worthy of comment? content and comment volume in political blogs. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, Washington, DC, May 2010. **ICWSM 2010.**
- [140] Shay B. Cohen and Noah A. Smith. Variational inference for grammar induction with prior knowledge. In *Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, companion volume*, pages 1–4, Singapore, August 2009. **ACL 2009.**
- [141] Sourish Chaudhuri, Naman K. Gupta, Noah A. Smith, and Carolyn P. Rosé. Leveraging structural relations for fluent compressions at multiple compression rates. In *Proceedings of the Joint Conference of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, companion volume*, pages 101–104, Singapore, August 2009. **ACL 2009.**
- [142] Ramnath Balasubramanyan, Frank Lin, William W. Cohen, Matthew Hurst, and Noah A. Smith. From episodes to sagas: Understanding the news by identifying temporally related story sequences. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*, San Jose, CA, May 2009. **ICWSM 2009.**
- [143] Daniel R. Rashid and Noah A. Smith. Relative keyboard input system. In *Proceedings of the International Conference on Intelligent User Interfaces*, pages 397–400, Canary Islands, Spain, January 2008. **IUI 2008.**
- [144] Markus Dreyer, David A. Smith, and Noah A. Smith. Vine parsing and minimum risk reranking for speed and precision. In *Proceedings of the Conference on Natural Language Learning*, pages 201–205, New York, NY, June 2006. **CoNLL 2006.**
- [145] Jason Eisner, Eric Goldlust, and Noah A. Smith. Dyna: A declarative language for implementing dynamic programs. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, companion volume*, pages 218–221, Barcelona, Spain, July 2004. **ACL 2004.**
- [146] Noah A. Smith and Michael E. Jahr. Cairo: An alignment visualization tool. In *Proceedings of the*

Language Resources and Evaluation Conference, pages 549–552, Athens, Greece, May/June 2000. **LREC 2000**.

Refereed workshop publications

- [147] Aaron Jaech, George Mulcaire, Mari Ostendorf, and Noah A. Smith. A neural model for language identification in code-switched tweets. In *Proceedings of the EMNLP Workshop on Computational Approaches to Linguistic Code Switching*, Austin, TX, November 2016. **LICS 2016**.
- [148] Aaron Jaech, George Mulcaire, Shobhit Hathi, Mari Ostendorf, and Noah A. Smith. Hierarchical character-word models for language identification. In *Proceedings of the International Workshop on Natural Language Processing for Social Media*, Austin, TX, November 2016. **SocialNLP 2016**.
- [149] Dani Yogatama, Bryan R. Routledge, and Noah A. Smith. A sparse and adaptive prior for time-dependent model parameters. In *Proceedings of the NIPS Workshop on Time Series*, Montréal, Québec, December 2015.
- [150] Sam Thomson, Brendan O’Connor, Jeffrey Flanigan, David Bamman, Jesse Dodge, Swabha Swayamdipta, Nathan Schneider, Chris Dyer, and Noah A. Smith. CMU: Arc-factored, discriminative semantic dependency parsing. In *Proceedings of the International (COLING) Workshop on Semantic Evaluations*, Dublin, Ireland, August 2014. **SemEval 2014**.
- [151] Shiladitya Sinha, Chris Dyer, Kevin Gimpel, and Noah A. Smith. Predicting the NFL using Twitter. In *Proceedings of the ECML/PKDD Workshop on (Machine Learning and Data Mining for) Sports Analytics*, Prague, Czech Republic, September 2013.
- [152] Nathan Schneider, Brendan O’Connor, Naomi Saphra, David Bamman, Manaal Faruqui, Noah A. Smith, Chris Dyer, and Jason Baldridge. A framework for (under)specifying dependency syntax without overloading annotators. In *Proceedings of the ACL Linguistic Annotation Workshop*, Sofia, Bulgaria, August 2013. **LAW 2013**.
- [153] Nathan Schneider, Chris Dyer, and Noah A. Smith. Exploiting and expanding corpus resources for frame-semantic parsing. April 2013. International FrameNet Workshop.
- [154] Jacob Eisenstein, Brendan O’Connor, Noah A. Smith, and Eric P. Xing. Mapping the geographical diffusion of new words. In *Proceedings of the NIPS Workshop on Social Network and Social Media Analysis: Methods, Models and Applications*, Lake Tahoe, NV, December 2012.
- [155] Waleed Ammar, Chris Dyer, and Noah A. Smith. Transliteration by sequence labeling with lattice encodings and reranking. In *Proceedings of the ACL Named Entities Workshop*, Jeju, Korea, July 2012.
- [156] Yanchuan Sim, Noah A. Smith, and David A. Smith. Discovering factions in the computational linguistics community. In *Proceedings of the ACL Workshop on Rediscovering Fifty Years of Discoveries*, Jeju, Korea, July 2012.
- [157] Brendan O’Connor, David Bamman, and Noah A. Smith. Computational text analysis for social science: Model complexity and assumptions. In *Proceedings of the NIPS Workshop on Computational Social Science and the Wisdom of Crowds*, Sierra Nevada, Spain, December 2011.
- [158] Chris Dyer, Kevin Gimpel, Jonathan H. Clark, and Noah A. Smith. The CMU-ARK German-English translation system. In *Proceedings of the EMNLP Workshop on Statistical Machine Translation*, Edinburgh, UK, July 2011. **SMT 2011**.
- [159] Kevin Gimpel and Noah A. Smith. Generative models of monolingual and bilingual gappy patterns. In *Proceedings of the EMNLP Workshop on Statistical Machine Translation*, Edinburgh, UK, July 2011. **SMT 2011**.
- [160] Desai Chen, Chris Dyer, Shay B. Cohen, and Noah A. Smith. Unsupervised bilingual POS tagging with Markov random fields. In *Proceedings of the EMNLP Workshop on Unsupervised Learning in*

- NLP*, Edinburgh, UK, July 2011. **UNSUP 2011.**
- [161] Jacob Eisenstein, Tae Yano, William W. Cohen, Noah A. Smith, and Eric P. Xing. Structured databases of named entities from Bayesian nonparametrics. In *Proceedings of the EMNLP Workshop on Unsupervised Learning in NLP*, Edinburgh, UK, July 2011. **UNSUP 2011.**
- [162] Dong Nguyen, Noah A. Smith, and Carolyn P. Rosé. Author age prediction from text using linear regression. In *Proceedings of the ACL Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, Portland, OR, June 2011. **LATECH 2011.**
- [163] André F. T. Martins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Augmenting dual decomposition for MAP inference. In *Proceedings of the International Workshop on Optimization for Machine Learning*, Whistler, BC, December 2010. **OPT 2010.**
- [164] André F. T. Martins, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Online multiple kernel learning for structured prediction. In *Proceedings of the NIPS Workshop on New Directions in Multiple Kernel Learning*, Whistler, BC, December 2010.
- [165] Brendan O’Connor, Jacob Eisenstein, Eric P. Xing, and Noah A. Smith. Discovering demographic language variation. In *Proceedings of the NIPS Workshop on Machine Learning for Social Computing*, Whistler, BC, December 2010.
- [166] Desai Chen, Nathan Schneider, Dipanjan Das, and Noah A. Smith. SEMAFOR: Frame argument resolution with log-linear models. In *Proceedings of the International (ACL) Workshop on Semantic Evaluations*, Uppsala, Sweden, July 2010. **SemEval 2010.**
- [167] Michael Heilman and Noah A. Smith. Extracting simplified statements for factual question generation. In *Proceedings of the AIED Workshop on Question Generation*, Pittsburgh, PA, June 2010.
- [168] Michael Heilman and Noah A. Smith. Rating computer-generated questions with Mechanical Turk. In *Proceedings of the NAACL-HLT Workshop on Creating Speech and Language Data With Mechanical Turk*, Los Angeles, CA, June 2010.
- [169] Tae Yano, Philip Resnik, and Noah A. Smith. Shedding (a thousand points of) light on biased language. In *Proceedings of the NAACL-HLT Workshop on Creating Speech and Language Data With Mechanical Turk*, Los Angeles, CA, June 2010.
- [170] Michael Heilman and Noah A. Smith. Ranking automatically generated questions as a shared task. In *Proceedings of the AIED Workshop on Question Generation*, Brighton, UK, July 2009.
- [171] André F. T. Martins and Noah A. Smith. Summarization with a joint model for sentence extraction and compression. In *Proceedings of the NAACL-HLT Workshop on Integer Linear Programming for Natural Language Processing*, Boulder, CO, June 2009.
- [172] Shay B. Cohen and Noah A. Smith. The shared logistic normal distribution for grammar induction. In *Proceedings of the NIPS Workshop on Speech and Language: Unsupervised Latent-Variable Models*, Whistler, BC, December 2008.
- [173] Noah A. Smith, Michael Heilman, and Rebecca Hwa. Question generation as a competitive undergraduate course project. In *Proceedings of the NSF Workshop on the Question Generation Shared Task and Evaluation Challenge*, Arlington, VA, September 2008.
- [174] Jason Eisner and Noah A. Smith. Competitive grammar writing. In *Proceedings of the ACL Workshop on Issues in Teaching Computational Linguistics*, pages 97–105, Columbus, OH, June 2008.
- [175] Kevin Gimpel and Noah A. Smith. Rich source-side context for statistical machine translation. In *Proceedings of the ACL Workshop on Statistical Machine Translation*, pages 9–17, Columbus, OH, June 2008. **SMT 2008.**
- Five-year retrospective best paper award.**
- [176] Noah A. Smith and Jason Eisner. Guiding unsupervised grammar induction using contrastive estimation. In *Proceedings of the IJCAI Workshop on Grammatical Inference Applications*, pages 73–82, Edinburgh, UK, July 2005.

Theses

- [177] Noah A. Smith. *Novel Estimation Methods for Unsupervised Discovery of Latent Structure in Natural Language Text*. Ph.D. thesis, Department of Computer Science, Johns Hopkins University, Baltimore, MD, October 2006. Supervised by Jason Eisner.
- [178] Noah A. Smith. Detection of translational equivalence. Technical report 4253, Department of Computer Science, University of Maryland College Park, College Park, MD, May 2001. Undergraduate honors thesis, supervised by Philip Resnik.
- [179] Noah A. Smith. Ellipsis happens, and deletion is how. In Andrea Gualmini, Soo-Min Hong, and Mitsue Motomura, editors, *University of Maryland Working Papers in Linguistics*, volume 11, pages 176–191. Department of Linguistics, University of Maryland, November 2001. Undergraduate honors thesis, supervised by Norbert Hornstein.
- [180] Alison J. Deming, Steven P. Denny, Jessica Exelbert, Anne Italiano, Dan Malinow, Katie E. Praske, Bradley Rhoderick, Noah A. Smith, Amanda Stamper, and Margaret E. Wood. Smart growth: An analysis in three counties, May 2001. University of Maryland College Park, Gemstone team thesis, supervised by Jacqueline Rogers.

OTHER RESEARCH OUTCOMES

Supervised doctoral theses

- [181] Yanchuan Sim. *Text as Strategic Choice*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, June 2016.
- [182] Waleed Ammar. *Towards a Universal Analyzer of Natural Languages*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, June 2016.
- [183] David Bamman. *People-Centric Natural Language Processing*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, July 2015.
Awarded Honorable Mention for the SCS Dissertation Prize.
- [184] Dani Yogatama. *Sparse Models of Natural Language Text*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, May 2015.
- [185] Brendan O’Connor. *Statistical Text Analysis for Social Science*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, August 2014.
- [186] Nathan Schneider. *Lexical Semantic Analysis in Natural Language Text*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, June 2014.
- [187] Tae Yano. *Text as Actuator: Text-Driven Response Modeling and Prediction in Politics*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, July 2013.
- [188] Kevin Gimpel. *Discriminative Feature-Rich Modeling for Syntax-Based Machine Translation*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, August 2012.
- [189] André F. T. Martins. *The Geometry of Constrained Structured Prediction: Applications to Inference and Learning of Natural Language Syntax*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, May 2012.
Awarded the IBM Portugal Prémio Científico and Honorable Mention for the SCS Dissertation Prize.
- [190] Dipanjan Das. *Semi-Supervised and Latent-Variable Models of Natural Language Semantics*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, April 2012.
- [191] Shay B. Cohen. *Computational Learning of Probabilistic Grammars in the Unsupervised Setting*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, September 2011.

- [192] Michael Heilman. *Automatic Factual Question Generation from Text*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, April 2011.

Supervised undergraduate theses

- [193] Desai Chen. Unsupervised bilingual POS tagging with Markov random fields, May 2011. Computer Science honors thesis, School of Computer Science, Carnegie Mellon University.
- [194] Rishav Bhowmick. [Rich entity type recognition in text](#), May 2010. Computer Science honors thesis, Department of Computer Science, Carnegie Mellon University–Qatar.
- [195] Neel Shah. [Predicting risk from financial reports with supervised topic models](#), May 2010. Computer Science honors thesis, School of Computer Science, Carnegie Mellon University.
- [196] Matthew Thompson. [Möbius: Exploring a new modality for poetry generation](#), April 2009. Linguistics senior thesis, Department of Philosophy, Carnegie Mellon University.

Unpublished technical reports and working papers

- [197] Dallas Card, Amber E. Boydston, and Noah A. Smith. Automated text analysis to unframe the news, June 2016. Presented at the 2nd International Conference on Computational Social Science.
- [198] Yanchuan Sim, Bryan R. Routledge, and Noah A. Smith. Text reveals influence: The case of amicus briefs and the supreme court, June 2016. Presented at the 2nd International Conference on Computational Social Science.
- [199] Philip Massey, Patrick Xia, David Bamman, and Noah A. Smith. Annotating character relationships in literary texts, December 2015.
- [200] Jason B. Colditz, Maharsi Naidu, Noah A. Smith, Joel Welling, and Brian A. Primack. Use of Twitter to assess sentiment toward waterpipe tobacco smoking, March 2016. Presented at the 37th Annual Meeting and Scientific Sessions of the Society of Behavioral Medicine.
- [201] Noah A. Smith, Claire Cardie, Anne L. Washington, and John D. Wilkerson. Overview of the 2014 NLP unshared task in PoliInformatics. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 5–7, Baltimore, MD, June 2014.
- [202] Amber E. Boydston, Dallas Card, Justin H. Gross, Philip Resnik, and Noah A. Smith. Tracking the development of media frames within and across policy issues, August 2014.
- [203] Lingpeng Kong and Noah A. Smith. An empirical comparison of parsing methods for stanford dependencies, April 2014.
- [204] Justin H. Gross, Brice Acree, Yanchuan Sim, and Noah A. Smith. Testing the etch-a-sketch hypothesis: A computational analysis of Mitt Romney’s ideological makeover during the 2012 primary vs. general elections, August 2013. Presented at the Annual Meeting of the American Political Science Association.
- [205] David Bamman and Noah A. Smith. New alignment methods for discriminative summarization, May 2013.
- [206] David Bamman, Adam Anderson, and Noah A. Smith. [Inferring social rank in an Old Assyrian trade network](#). July 2013. Presented at Digital Humanities.
- [207] Waleed Ammar, Shomir Wilson, Norman Sadeh, and Noah A. Smith. Automatic categorization of privacy policies: A pilot study. Technical Report CMU-LTI-12-019, Carnegie Mellon University, Pittsburgh, PA, December 2012.
- [208] Noah A. Smith. Adversarial evaluation for models of natural language, July 2012.
- [209] Chris Dyer, Noah A. Smith, Graham Morehead, Phil Blunsom, and Abby Levenberg. The CMU-Oxford translation system for the NIST open machine translation 2012 evaluation, May 2012.

- [210] Tal Stramer, Bryan R. Routledge, and Noah A. Smith. [Predicting FED action from text](#). Technical Report CMU-LTI-11-005, Carnegie Mellon University, Pittsburgh, PA, May 2011.
- [211] Nathan Schneider, Rebecca Hwa, Philip Gianfortoni, Dipanjan Das, Michael Heilman, Alan W. Black, Frederick L. Crabbe, and Noah A. Smith. Visualizing topical quotations over time to understand news discourse. Technical Report CMU-LTI-10-013, Carnegie Mellon University, Pittsburgh, PA, July 2010.
- [212] André F. T. Martins, Kevin Gimpel, Noah A. Smith, Eric P. Xing, Pedro M. Q. Aguiar, and Mário A. T. Figueiredo. Aggressive online learning of structured classifiers. Technical Report CMU-ML-10-109, Carnegie Mellon University, Pittsburgh, PA, June 2010.
- [213] Kevin Gimpel and Noah A. Smith. Softmax-margin training for structured log-linear models. Technical Report CMU-LTI-10-008, Carnegie Mellon University, Pittsburgh, PA, June 2010.
- [214] Noah A. Smith. Text-driven forecasting, March 2010.
- [215] Michael Heilman and Noah A. Smith. Question generation via overgenerating transformations and ranking. Technical Report CMU-LTI-09-013, Carnegie Mellon University, Pittsburgh, PA, June 2009.
- [216] Dan Tasse and Noah A. Smith. SOUR CREAM: Toward semantic processing of recipes. Technical Report CMU-LTI-08-005, Carnegie Mellon University, Pittsburgh, PA, May 2008.
- [217] Yaser Al-Onaizan, Jan Čuřin, Michael Jahr, Kevin Knight, John Lafferty, I. Dan Melamed, Noah A. Smith, Franz-Josef Och, David Purdy, and David Yarowsky. Statistical machine translation. CLSP Research Notes 42, Johns Hopkins University, Baltimore, MD, 1999.
- [218] Margaret E. Wood, Noah A. Smith, Anne Italiano, Jessica Exelbert, Steven Denny, and Konrad Aschenbach. Edgewood Terrace: The decline and revitalization of a mixed-income neighborhood. November 1999. Association for Public Policy Analysis and Management Fall Conference on Global and Comparative Perspectives.

Publicly available datasets and software

- [219] [CHARACTERRELATIONS](#), developed by Philip Massey, Patrick Xia, and David Bamman. Annotations of character relations in 109 literary texts, 2015, see [199].
- [220] [GFL-WEB](#), developed by Michael T. Mordowanec. Web interface for the Graph Fragment Language for syntactic annotation, 2014, see [128].
- [221] [JAMR](#), developed by Jeffrey Flanigan and Sam Thomson. Semantic parser for the Abstract Meaning Representation, 2014, see [53].
- [222] [COMPREHENSIVE MULTIWORD EXPRESSIONS CORPUS](#), developed by Nathan Schneider. English Web Treebank annotated with multiword expressions, 2014, see [56].
- [223] [AMALGR](#), developed by Nathan Schneider. Multiword expression identification tool, 2014, see [11].
- [224] [NFL TWEET DATASET](#), developed by Shiladitya Sinha, Brendan O’Connor, Chris Dyer, and Kevin Gimpel. NFL game data and identifiers of tweets aligned to games, 2013, see [151].
- [225] [CMU MOVIE SUMMARY CORPUS](#), developed by David Bamman and Brendan O’Connor. Corpus of movie plot summaries and associated metadata, 2013, see [60].
- [226] [GLOBAL VOICES MALAGASY-ENGLISH PARALLEL CORPUS](#), developed by Victor Chahuneau. Corpus of parallel news articles from the Global Voices citizen media project, 2012.
- [227] [WORD SALAD](#), developed by Victor Chahuneau. Corpus of restaurant menus, 2012, see [64].
- [228] [ARABIC WIKIPEDIA SUPERSENSE CORPUS](#), developed by Nathan Schneider and Behrang Mohit. Articles tagged with nominal supersenses, 2012, see [133].
- [229] [AD³](#), developed by André Martins. Approximate MAP decoder, 2012, see [6, 66, 73, 76, 164].

- [230] [RAMPION](#), developed by Kevin Gimpel. Algorithm for training statistical machine translation models based on minimizing structured ramp loss, 2012, see [68].
- [231] [CONGRESSIONAL BILLS CORPUS](#), developed by Tae Yano. Congressional bills and committee outcomes, 2012, see [69].
- [232] [ARABIC WIKIPEDIA NAMED ENTITY CORPUS AND TAGGER](#), developed by Nathan Schneider and Behrang Mohit. Articles tagged with named entities, and a statistical tagger, 2012, see [70].
- [233] [TWITTER PART-OF-SPEECH TAGGING](#), developed by Olutobi Owoputi, Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan, Chris Dyer, and Noah A. Smith. Dataset of tweets manually annotated with part-of-speech tags; part-of-speech tagger trained on this data; a simple browser-based POS tagging annotation interface, 2011, see [63, 136].
- [234] [QUESTION-ANSWER DATA](#), developed by Michael Heilman, Shay Cohen, and Kevin Gimpel. Question-answer pairs generated by undergraduates for the purpose of developing and evaluating question answering systems., 2010, see [173].
- [235] [AMAZON MECHANICAL TURK POLITICAL BIAS DATA](#), developed by Tae Yano. Sentences from political blogs with crowdsourced annotations of political bias, 2010, see [169].
- [236] [MOVIE\\$ CORPUS](#), developed by Mahesh Joshi, Dipanjan Das, and Kevin Gimpel. Collection of pre-release movie reviews, metadata, and opening weekend revenues, 2010, see [138].
- [237] [SEMAFOR](#), developed by Dipanjan Das, Nathan Schneider, and Desai Chen. Frame-semantic parser, 2010, see [88].
- [238] [TURBOPARSER](#), developed by André Martins. Multilingual dependency parser, 2009, see [94].
- [239] [QUIPU](#), developed by Kevin Gimpel. Statistical machine translation system, 2009, see [92].
- [240] [POLITICAL BLOG CORPUS](#), developed by Tae Yano. Text collection from five American political blogs, 2009, see [99].
- [241] [10-K CORPUS](#), developed with three others. Collection of 10-K reports and preceding and following stock return volatility measurements, 2009, see [97].
- [242] [DAGEEM](#), developed by Shay Cohen. Unsupervised dependency grammar induction, 2008, see [102].
- [243] [MSTPARSER, STACKED](#), developed by André Martins and Dipanjan Das. Multilingual dependency parser, 2008, see [103].
- [244] [CURD](#), developed by Dan Tasse. Corpus of semantically annotated recipes, 2008, see [216].
- [245] [DYNA](#), developed with five others. Declarative programming language for weighted dynamic programming, 2004, see [113, 145].
- [246] [EGYPT](#), developed with nine others. Toolkit for statistical machine translation, including GIZA training module and CAIRO word alignment visualizer, 1999, see [146, 217].