

## Noah A. Smith

@ [nasmith@cs.washington.edu](mailto:nasmith@cs.washington.edu) @ [noah@allenai.org](mailto:noah@allenai.org)  
✉ Box 352350, University of Washington ✉ 2157 N. Northlake Way, Suite 110  
Seattle, WA 98195 Seattle, WA 98103  
☎ +1-206-685-3134  
ORCID 0000-0002-2310-6380  
🏠 [nasmith.github.io](https://nasmith.github.io)

### PROFESSIONAL EXPERIENCE

University of Washington	2021– 2018– 2015–2018	<b>Amazon Professor of Machine Learning Professor</b> <b>Associate Professor</b> Paul G. Allen School of Computer Science and Engineering Senior Data Science Fellow, eScience Institute Affiliate, Center for Statistics and the Social Sciences Adjunct, Linguistics
Allen Institute for AI	2018–	<b>Senior Director of NLP Research</b>
Carnegie Mellon University	2011–2015 2006–2011	<b>Associate Professor</b> (tenure awarded 7/1/2014) <b>Assistant Professor</b> 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	<b>Ph.D.</b> , Computer Science <i>Thesis topic:</i> unsupervised natural language parsing [304]
	2004	<b>M.S. in Engineering</b> , Computer Science
University of Maryland	2001	<b>B.S.</b> with High Honors in Computer Science
	2001	<b>B.A.</b> with Honors in Linguistics <i>Summa cum laude; honors theses:</i> [305–307]
University of Edinburgh	2000	Visiting student in Linguistics and Artificial Intelligence

## SELECTED HONORS AND AWARDS

- Fellow of the Association for Computational Linguistics (elected 2020, “for significant contributions to linguistic structure prediction, computational social sciences, and improving NLP research methodology”)
- Inaugural Amazon Alexa prize first-place team co-mentor with M. Ostendorf and Y. Choi (2017)
- UW innovation award “to stimulate innovation among faculty from a range of disciplines and to reward some of their most terrific ideas” (2016–2018)
- CMU Finmeccanica career development chair, endowed “to acknowledge promising teaching and research potential in junior faculty members” (2011–2014)
- Various papers awarded best/outstanding paper [52, 66, 112, 122, 139, 192, 199, 268, 302] or finalist/honorable mention/nominated [84, 108, 131, 151, 160, 182, 206, 212, 228, 229, 265]
- Fannie and John Hertz Foundation fellowship (2001–2006)

## SELECTED PRESS

- [These simple changes can make AI research much more energy efficient.](#) *MIT Technology Review*. July 6, 2022. [55]
- [It began as an AI-fueled dungeon game. It got much darker.](#) *Wired*. May 5, 2021. [77]
- [Your favorite A.I. language tool is toxic.](#) *Fortune*. September 29, 2020. [77]
- [Would Elizabeth Warren or Bernie Sanders move to the center against Trump?](#) *FiveThirtyEight*. October 1, 2019. [13]
- [At tech’s leading edge, worry about a concentration of power.](#) *New York Times*. September 26, 2019. [10]
- [Artificial intelligence confronts a ‘reproducibility’ crisis.](#) *Wired*. September 16, 2019. [92]
- [Google’s hate speech-detecting AI appears to be racially biased.](#) *New Scientist*. August 12, 2019. [228]
- [AI researchers need to stop hiding the climate toll of their work.](#) *MIT Technology Review*. August 9, 2019. [10]
- [How to teach artificial intelligence some common sense.](#) *Wired*. November 13, 2018. [100]
- [Inside the Alexa Prize.](#) *Wired*. February 27, 2018. [233]
- [10 inventions predicted by \*The Simpsons\*.](#) *Alltime10s*. May 5, 2016. [243]
- [My favorite thing about the Internet? Definitely the sarcasm.](#) *MIT Technology Review*. January 21, 2016. [243]
- [Talk to me.](#) *New Yorker*. February 9, 2015. [180]
- [The machines vs. Mitt Romney: How artificial intelligence is parsing political rhetoric.](#) *Popular Science’s Zero Moment*. September 24, 2014. [157, 368]
- [Addiction and seduction on Yelp: The language of food love.](#) *NPR’s All Things Considered*. April 16, 2014. [25]
- [The linguist’s mother lode: What Twitter reveals about slang, gender, and no-nose emoticons.](#) *Time*. September 9, 2013. [180, 189, 281]
- [Do you have a Twitter ‘accent’?](#) *NPR’s Here and Now*. September 4, 2013. [281]
- [Main tweet: Researchers dig into the intersection of politics and Twitter.](#) *Time’s Swampblog*. August 14, 2013. [189, 248]

- [Here's how you can use Twitter to beat the spread on sports betting.](#) *Washington Post*, August 12, 2013. [278]
- [Congress' magic words.](#) *Washington Post's Wonkblog*. December 31, 2012. [167]
- [China's social networks hit by censorship, says study.](#) *BBC*. March 9, 2012. [32]
- [Revealed: How China censors its social networks.](#) *New Scientist*. March 8, 2012. [32]
- [Twitterology: A new science?](#) *New York Times*. October 30, 2011. [180]
- [Noah Smith on supercomputer trivia and Twitter dialects.](#) *CBC's Spark*. February 15, 2011. [180]
- [You have an accent even on Twitter.](#) *NPR's All Things Considered*. January 18, 2011. [180]
- [Twitter a decent stand-in for public opinion polls.](#) *Ars Technica*. May 11, 2010. [189]

## RESEARCH FUNDING

### *Current sponsored research (\*lead PI)*

NSF	*Efficient transformers via formal and empirical analysis (with R. Schwartz)	\$500K	2021–24
Allstate	History-informed dialog system for customer interaction (with M. Ostendorf)	\$1.39M	2019–22
ONR	Generating documents that are consistent with a knowledge base (with S. Jajodia, H. Hajishirzi, P. Resnik, and V.S. Subrahmanian)	\$1.89M	2018–23

### *Industry gifts, internal, and other*

Amazon	Language model customization (includes AWS credit)	\$163K	2021
Apple	Customizable multilingual language models	\$99K	2021
T-Mobile	Emerging topic and outlier detection (with M. Ostendorf)	\$237K	2019
Microsoft	Modeling diversity in the machine learning pipeline (Azure credit)	\$20K	2017
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	Innovation award: 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)		2011–8
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	Civil rights violations and the democratic rule of law (PI: E. Gade)	\$249K	2018–21
NSF	*Modeling crosslinguistic influences between language varieties (with N. Schneider, Y. Tsvetkov, and S. Wintner)	\$167K	2018–21
NSF	*Broad-coverage semantic parsing: linguistic representation learning from crowd-scale data (with L. Zettlemoyer)	\$1.18M	2016–21
DARPA	Communicating intelligently with computers (Communicating with Computers program; subcontract to ISI, UW PI: Y. Choi)	\$1.99M	2015–20
NSF	Simulation-based hypothesis testing of socio-technical community resilience using distributed optimization and natural language processing (PI: S. Miles)	\$1.21M	2015–21
Allstate	Prediction from text with explanations	\$293K	2017–8
IARPA	Foreign language automated information retrieval (Machine Translation for English Retrieval of Information in Any Language program; subcontract to BBN)	\$118K	2017–8
DARPA	Language technology for rapid response (Low Resource Languages for Emergent Incidents program; subcontract to BBN)	\$480K	2015–8
DARPA	Analysis of rare incident-event languages (Low Resource Languages for Emergent Incidents program; subcontract to CMU)	\$84K	2015–17
DARPA	Structured distributed semantics: analysis and filtering of text (Deep Exploration and Filtering of Text program; PI: E. Hovy)	\$5.66M	2012–17
NSF	Towards effective web privacy notice and choice: a multi-disciplinary perspective (PI: N. Sadeh; 1330596)	\$3.75M	2013–17
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. Boydstun, 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](#) (2011–21)
- [Structured Prediction for Natural Language Processing](#)  
Three-hour invited tutorial at ICML 2009

#### *Graduate courses*

- [Natural Language Processing](#) Spring 2017  
Professional masters program course
- [Natural Language Processing](#) Winter 2016, Spring 2018, Winter 2020–2  
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 courses*

- Machine Learning Autumn 2017  
Broad introduction to the field; Daumé textbook
- Natural Language Processing Spring 2008–11, 2013–4; Winter 2017, 2019, 2021, 2022  
Broad introduction to the field (designer & instructor)  
Co-taught with Chris Dyer in 2014
- Natural Language Processing Capstone Spring 2019–22  
Project-based course

#### *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 [301], taught by others since 2004
- [Introduction to Programming for Linguists](#) (teaching assistant) Fall 2000  
Department of Linguistics, University of Maryland

## ADVISING

### *Ph.D., completed*

1. [Michael Heilman](#)<sup>1</sup> (2008–2011); NSF graduate fellow, PIER scholar; [173, 187, 188, 254, 294, 295, 297, 300, 330, 377, 381, 407]; research scientist at Educational Testing Service; data scientist at Civis Analytics
2. [Shay B. Cohen](#)<sup>2</sup> (2006–2011); ICLP 2008 best student paper; Computing Innovation postdoctoral fellowship; [31, 33, 35, 169, 179, 184, 185, 194, 199, 200, 204, 258, 287, 299, 329, 416]; postdoctoral fellow at Columbia University, then lecturer/reader ( $\approx$  assistant/associate professor) at the University of Edinburgh
3. [Dipanjan Das](#)<sup>3</sup> (2008–2012); ACL 2011 best paper; [28, 164, 165, 169, 175, 183, 186, 191, 201, 254, 256, 293, 328, 377, 403, 407, 410, 417]; research scientist at Google
4. [André F. T. Martins](#)<sup>4</sup> (2007–2012); ICTI/Portugal scholar; ACL 2009 best paper; SCS Dissertation Award Honorable Mention; IBM Portugal *Prémio Científico*; [21, 28, 36, 164, 171, 172, 174, 178, 181, 192, 193, 201, 203, 290, 291, 298, 327, 378, 412, 417]; research scientist at Priberam, then Unbabel; VP of AI Research at Unbabel and associate professor at Instituto Superior Técnico
5. [Kevin Gimpel](#)<sup>5</sup> (2006–2012); Sandia-CMU graduate fellowship; [24, 162, 166, 170, 183, 190, 198, 200, 252, 254–256, 285, 286, 302, 326, 404, 407, 410, 413]; assistant research professor, then assistant professor at TTI-Chicago
6. [Tae Yano](#)<sup>6</sup> (2007–2013); [167, 197, 248, 257, 288, 296, 325, 405, 414]; research engineer at Microsoft, then research scientist at Expedia
7. [Nathan Schneider](#)<sup>7</sup> (2008–2014); [26, 28, 147, 154, 161, 168, 186, 249, 251, 254, 277, 279, 280, 293, 324, 377, 396, 397, 402, 406, 407]; postdoctoral researcher at the University of Edinburgh then assistant/associate professor at Georgetown University
8. [Brendan O'Connor](#)<sup>8</sup> (2009–2014); [22, 32, 158, 159, 161, 173, 180, 189, 254, 277, 279, 281, 284, 292, 323, 398, 399, 407]; assistant/associate professor at the University of Massachusetts Amherst
9. [Dani Yogatama](#)<sup>9</sup> (2010–2015); [27, 137, 138, 141, 152, 153, 163, 173, 239, 240, 248, 254, 275, 322, 407]; research scientist at Baidu, then Deepmind, then assistant professor at the University of Southern California

<sup>1</sup>Began Ph.D. in 2007. Proposal: September 1, 2009. Defense: March 23, 2011.

<sup>2</sup>Proposal: September 8, 2010. Defense: September 27, 2011.

<sup>3</sup>Proposal: April 23, 2010. Defense: April 23, 2012.

<sup>4</sup>Co-advisors: Eric Xing (CMU), Mário Figueiredo and Pedro Aguiar (IST, Universidade Técnica de Lisboa). Proposal: August 25, 2009. Defense: May 11, 2012. Began Ph.D. in 2006, dual CMU-Portugal program in 2007.

<sup>5</sup>Proposal: June 15, 2010. Defense: August 28, 2012.

<sup>6</sup>Co-advisor: William Cohen (CMU). Proposal: September 14, 2011. Defense: July 2, 2013

<sup>7</sup>Proposal: November 28, 2012. Defense: June 19, 2014.

<sup>8</sup>Began as M.S. student in 2009, joined Ph.D. program in 2010. Proposal: April 9, 2013. Defense: August 19, 2014.

<sup>9</sup>Began as M.S. student in 2010, joined Ph.D. program in 2012. Proposal: April 16, 2014. Defense: May 27, 2015.



10. [David Bamman](#)<sup>10</sup> (2011–2015); Alan J. Perlis SCS graduate student teaching award; SCS Dissertation Award Honorable Mention; [32, 133, 150, 158, 243, 244, 277, 279, 284, 321, 371, 372, 399]; assistant/associate professor at the University of California Berkeley
11. [Waleed Ammar](#)<sup>11</sup> (2011–2016); Google fellowship in natural language understanding; [20, 146, 282, 320, 364, 373]; research scientist at the Allen Institute for Artificial Intelligence, then Google
12. [Yanchuan Sim](#)<sup>12</sup> (2011–2016); A\*STAR national science scholarship (Singapore); [13, 134, 143, 145, 156, 157, 163, 283, 319, 370]; research scientist at Institute for Infocomm Research; co-founder and CTO of Bot M.D.
13. [Lingpeng Kong](#)<sup>13</sup> (2013–2017); [16, 122, 124, 130, 140, 147, 238, 318, 369]; research scientist at Deepmind, then assistant professor at Hong Kong University
14. [Jeffrey Flanigan](#)<sup>14</sup> (2014–2018); [129, 141, 151, 254, 277, 317, 395, 407]; postdoctoral researcher at University of Massachusetts Amherst then assistant professor at the University of California Santa Cruz
15. [Sam Thomson](#)<sup>15</sup> (2012–2019); [103, 104, 108, 110, 113, 120, 141, 151, 242, 277, 316, 362, 385, 395]; research scientist at Semantic Machines
16. [Swabha Swayamdipta](#)<sup>16</sup> (2013–2019); [104, 113, 127, 147, 232, 234, 277, 315, 357, 362]; young investigator at the Allen Institute for Artificial Intelligence, then assistant professor at the University of Southern California
17. [Dallas Card](#)<sup>17</sup> (2013–2019); NSERC postgraduate scholarship (Canada); [5, 12, 92, 97, 101, 107, 111, 121, 123, 228, 241, 314, 360, 368, 386, 387]; postdoctoral researcher at Stanford, then assistant professor at the University of Michigan
18. [Jesse Dodge](#)<sup>18</sup> (2013–2020); NAACL 2015 best student paper; [10, 89, 92, 139, 226, 277, 313, 355, 363]; research scientist at the Allen Institute for Artificial Intelligence
19. [Elizabeth Clark](#)<sup>19</sup> (2015–2021); NSF graduate fellow, 2017 Alexa prize; [66, 96, 112, 116, 221, 233, 266, 312]; research scientist at Google
20. [Maarten Sap](#)<sup>20</sup> (2015–2021); 2017 Alexa prize; [68, 77, 88, 100, 109, 118, 224, 228, 233, 266, 270, 311]; young investigator at the Allen Institute for Artificial Intelligence, then assistant professor at Carnegie Mellon University
21. [Lucy Lin](#)<sup>21</sup> (2015–2021); NSF graduate fellow; [106, 222, 310, 356, 358–360]; research scientist at Spotify
22. [Phoebe Mulcaire](#) (2015–2022)<sup>22</sup> [20, 76, 80, 91, 231, 232, 271, 272, 309, 364]; research scientist at Duolingo
23. [Hao Peng](#)<sup>23</sup> (2016–2022); Google fellowship in natural language processing; [41, 43, 58, 61, 72, 73, 86, 92, 103, 108, 113, 115, 120, 216, 227, 308]; young investigator at the Allen Institute for AI, then assistant professor at the University of Illinois Urbana-Champaign

<sup>10</sup>Proposal: August 5, 2014. Defense: July 13, 2015.

<sup>11</sup>Co-advisor: Chris Dyer (CMU). Proposal: October 2, 2014. Defense: June 8, 2016.

<sup>12</sup>Proposal: June 11, 2015. Defense: June 27, 2016.

<sup>13</sup>Co-advisor: Chris Dyer (CMU). Began as M.S. student in 2012, became my student in 2013, joined Ph.D. program in 2015. Proposal: October 4, 2016. Defense: September 18, 2017.

<sup>14</sup>Co-advisors: Chris Dyer (CMU) and Jaime Carbonell (CMU). Began as M.S. student in 2010, joined Ph.D. program in 2012, became my student in 2014. Proposal: April 7, 2016. Defense: November 29, 2018.

<sup>15</sup>Began as M.S. student in 2012, joined Ph.D. program in 2014. Proposal: May 2, 2018. Defense: February 27, 2019.

<sup>16</sup>Co-advisor: Chris Dyer (CMU). Proposal: September 19, 2017. Defense: May 8, 2019.

<sup>17</sup>Proposal: September 20, 2017. Defense: August 7, 2019.

<sup>18</sup>Proposal: February 23, 2018. Defense: May 8, 2020.

<sup>19</sup>General exam: December 2, 2020. Defense: June 24, 2021.

<sup>20</sup>Co-advisor: Yejin Choi. General exam: May 20, 2020. Defense: July 1, 2021.

<sup>21</sup>General exam: May 19, 2020. Defense: December 2, 2021.

<sup>22</sup>General exam: March 10, 2021. Defense: May 9, 2022.

<sup>23</sup>General exam: February 10, 2021. Defense: June 28, 2022.



24. Kelvin Luu<sup>24</sup> (2015–2022) [14, 53, 69]; research scientist at Duolingo
25. Tal August<sup>25</sup> (2018–2022) [56, 66, 75, 90, 102, 266]; young investigator at the Allen Institute for Artificial Intelligence

*Ph.D., all-but-dissertation (completed thesis proposal or general examination)*

- Ofir Press<sup>26</sup> (2018–) [59, 70, 87, 361]
- Jungo Kasai<sup>27</sup> (2018–); IBM Ph.D. fellowship; [41, 43, 44, 50–52, 58, 61, 62, 72, 91, 231, 353]

*Ph.D.*

- Sofia Serrano (2017–) [66, 98, 358]
- Nikita Haduong<sup>28</sup> (2018–) [66]
- Yizhong Wang<sup>29</sup> (2019–) [47, 62, 81]
- Phillip Keung (2020–) [9, 217, 218, 223]
- Suchin Gururangan<sup>30</sup> (2020–); Bloomberg data science fellowship; [40, 49, 53, 66, 219]
- Alisa Liu<sup>31</sup> (2020–); NSF graduate fellow; [45, 68]
- Rahul Nadkarni<sup>32</sup> (2020–) [65]
- Orevaoghene Ahia<sup>33</sup> (2021–)
- Yushi Hu<sup>34</sup> (2022–) [42, 46]

*Postdoctoral*

- Chris Dyer (2010–2); [30, 173, 176, 253, 282, 285, 287, 375]; assistant professor at CMU, then researcher at Deepmind
- Behrang Mohit (2010–2), co-supervised by Kemal Oflazer (CMU–Qatar); [168, 249, 251, 402, 406]; research scientist at Apple
- Fei Liu (2013–5), co-advised by Norman Sadeh; [131, 141, 148, 240, 245, 276]; assistant professor at the University of Central Florida, then associate professor at Emory University
- Chenhao Tan (2016–7) [107, 115–117, 121, 387]; assistant professor at the University of Colorado Boulder then University of Chicago
- Yangfeng Ji (2016–8) [112, 116, 117, 119]; assistant professor at the University of Virginia
- Roy Schwartz (2016–9) [10, 92, 95, 103, 110, 118, 226, 227, 230, 234, 268, 270, 385]; lecturer ( $\approx$  assistant professor) at the Hebrew University in Jerusalem
- Emily Kalah Gade (2017–9); assistant professor at Emory University [5]
- Rachel Rudinger (2019–20); assistant professor at the University of Maryland [82]

<sup>24</sup>General exam: December 7, 2021. Defense: July 20, 2022.

<sup>25</sup>Co-advisor: Katharina Reinecke. Began as Ph.D. student in 2017, added me as advisor in 2018. General exam: February 24, 2021. Defense: November 7, 2022.

<sup>26</sup>General exam: August 3, 2022.

<sup>27</sup>General exam: August 10, 2022.

<sup>28</sup>Co-advisor: Steve Seitz.

<sup>29</sup>Co-advisor: Hanna Hajishirzi.

<sup>30</sup>Co-advisor: Luke Zettlemoyer.

<sup>31</sup>Co-advisor: Yejin Choi.

<sup>32</sup>Began as Ph.D. student in 2015, became my advisee in 2020.

<sup>33</sup>Co-advisor: Yulia Tsvetkov.

<sup>34</sup>Co-advisor: Mari Ostendorf. Began as Ph.D. student in 2021, added me as advisor in 2022.

- [Rik Koncel-Kedziorski](#) (2019–21); research scientist at Amazon [57, 69]
- [Sarah K. Dreier](#) (2019–21); assistant professor at the University of New Mexico [5]
- Nikolaos Pappas (2019–21); research scientist at Amazon [58, 72, 73, 78, 80, 83, 220]
- Tao Yu (2021–2), coadvised by Mari Ostendorf; assistant professor at Hong Kong University [42, 48]
- Hila Gonen (2021–), coadvised by Luke Zettlemoyer
- Wenya Wang (2021–), coadvised by Hanna Hajishirzi; assistant professor at Nanyang Technological University, Singapore

*M.S.*

- [Cari \(Sisson\) Bader](#) (2007–8); engineer at Nuance
- [Daniel Mills](#) (2010–1); [254, 407]
- [Victor Chahuneau](#)<sup>35</sup> (2012–3) [25, 30, 155, 160, 162, 250, 400, 401]
- [Rohan Ramanath](#)<sup>36</sup> (2013–5) [131, 148, 245, 276]; engineer at LinkedIn
- Yi Zhu (2016–7) [114, 348]; Ph.D. student at Cambridge University
- Suchin Gururangan (2017–8) [234, 347]; predoctoral investigator at AI2, then Ph.D. student at UW
- Tam Dang (2018–9) [97, 341]; Xevo
- Ron Fan (2018–9) [342]; AI2 Incubator
- Aishwarya Nirmal (2018–9) [345]; Airbnb
- Deric Pang (2018–9) [346]; Google
- Sam Gehman (2019–20) [77, 339]; Stripe
- Karishma Mandyam (2019–20) [53, 340]; Facebook
- Xinyi Wu (2019–20) [69]; Amazon
- Kyle Yan (2019–21) [217]; Google
- Xuhui Zhou (2019–21) [74, 83, 337]; Ph.D. student at CMU
- Zhaofeng Wu (2020–1) [8, 336]; predoctoral investigator at AI2 then Ph.D. student at MIT
- Zeyu Blarry Wang (2020–1) [40, 335]
- Ethan Chau (2020–1) [222, 265, 334, 338]
- Zhilin Wang (2021) [333]
- Jacob Morrison (2021–2) [50, 51]; predoctoral investigator at AI2
- Connor Boyle (2021–) [48]
- Ivan Montero (2021–2) [41, 78, 220, 332]
- Tobias Rohde (2021–2) [331]
- Kaiser Sun (2021–2); Ph.D. student at Johns Hopkins University

*Student and postdoctoral visitors and independent study projects*

2018: Judit Ács (Budapest University of Technology and Economics), Qiaolin Xia (Peking U.) [94, 354];  
 2016: Adhi Kuncoro (CMU) [122, 124, 128], Yijia Liu (Harbin Institute of Technology) [114]; 2015: Miguel Ballesteros (Universitat Pompeu Fabra) [17, 122, 124, 127, 132, 136, 236], Renato Negrinho (Instituto Superior Técnico); 2014: Zita Marinho [125]; 2013: Daniel Preotiu-Pietro (Sheffield University), Lingpeng Kong; Minghui Qiu (Singapore Management University) [143, 156]; 2012: Swapna Gottipati

---

<sup>35</sup>Co-advisor: Chris Dyer (CMU).

<sup>36</sup>Co-advisor: Norman Sadeh (CMU).

(Singapore Management University) [156]; 2011: Dong Nguyen [289]; 2008: Aaron Phillips, Narges Sharif-Razavian, Sourish Chaudhuri [259], Severin Hacker, ThuyLinh Nguyen [182]; 2007: Daniel Rashid [261], Mengqiu Wang [206]; 2006: Thuy Linh Nguyen

#### *Ph.D. thesis committees*

*Ongoing*: Orion Dollar (UW Chemical Engineering), Agatha Downey (UW Linguistics), Ben Lee, Julian Michael, Woodley Packard (UW Linguistics), Francisco Santamarino (UW Evans School), Naomi Shapiro (UW Linguistics); 2022: Sarah Wiegrefe (Georgia Tech.), Tongshuang Sherry Wu, James Ferguson; 2021: Maxwell Forbes [57], John Thickstun; 2020: Hannah Rashkin [100, 109], Vicky Zayats (UW ECE), Antoine Bosselut; 2018: Yishu Miao (Oxford), Aaron Jaech (UW ECE) [271, 272], Luheng He, Nicholas FitzGerald; 2017: Adam Anderson (Harvard) [372], Kenton Lee; 2016: Manaal Faruqui (CMU) [137–139], Yangfeng Ji (Georgia Tech.), Yulia Tsvetkov (CMU) [137, 364]; 2015: Dan Garrette (U. Texas) [135, 144, 149], Jonathan Clark [176, 253], Jayant Krishnamurthy, Ankur Parikh, Wang Ling [136]; 2014: Anil Nelakanti (Université Pierre et Marie Curie); 2013: Khalid El-Arini, Mahesh Joshi [256], Ramnath Balasubramanyan [189, 260], Vladimir Eidelman (U. Maryland); 2012: Aaron Phillips, Sanjika Hewavitharana; 2011: Ming-Wei Chang (UIUC), David Huggins-Daines, Andreas Zollmann [196, 202]; 2010: Andrew Carlson; 2009: Andrew Arnold; 2008: Ashish Venugopal [196, 202], Ying Joy Zhang

#### *Undergraduate research advisees*

UW (excludes those who continued into fifth-year masters): Daniel Cheng [217], Lavinia Dunagan [50, 51], Ivy Guo, Skyler Hallinen, Mourad Heddaya,<sup>37</sup> Sean Hung, Chase Lee, Alan Li, Millicent Li,<sup>38</sup> Ryan Li, Zeyu Leo Liu [62],<sup>39</sup> Nelson Liu [99, 225, 230, 268, 269, 343],<sup>40</sup> Karishma Mandyam [344], Jacob Morrison, Kai Nylund, Nikko Rush, Amy Shah, Amol Sharma, Leroy Wang [40], Johnny Wei (UMass),<sup>41</sup> Stephen Wu, Velocity Yu, Michael Zhang [101],<sup>42</sup> Daniel Zhu  
CMU: Rishav Bhowmick (CMU–Qatar) [168, 350], Desai Chen [28, 186, 287, 293, 349],<sup>43</sup> Arash Enayati (CMU–Qatar), Mohammad Haque, Dimitry Levin [195], Philip Massey [365, 393], Zack McCord, Tobi Owoputi [161], Benjamin Plaut,<sup>44</sup> Naomi Saphra [279],<sup>45</sup> Neel Shah [351], Shiladitya Sinha [278, 398], Tal Stramer [376], Matthew Thompson [352], Daniel Tasse [382, 418],<sup>46</sup> Patrick Xia [365, 393],<sup>47</sup> Xiaote Zhu

#### *High school research advisees*

Katya Anderson, Lily Scherlis [162]

#### *Full-time technical staff*

- Zach Paine (2008); later at Google, Apple

---

<sup>37</sup>Pursuing a Ph.D. at U. Chicago.

<sup>38</sup>Pursuing a Ph.D. at Northeastern U.

<sup>39</sup>AI Resident at Facebook.

<sup>40</sup>Finalist, CRA Outstanding Undergraduate Research Award, 2019. Paul G. Allen School best senior thesis award, 2019. Pursuing a Ph.D. at Stanford.

<sup>41</sup>Pursuing a Ph.D. at U. Southern California.

<sup>42</sup>Pursuing a Ph.D. at U. Texas.

<sup>43</sup>Honorable mention, CRA Outstanding Undergraduate Research Award, 2010. Ph.D. at MIT.

<sup>44</sup>Pursuing a Ph.D. at Stanford.

<sup>45</sup>Pursuing a Ph.D. at U. Edinburgh.

<sup>46</sup>Ph.D. at CMU.

<sup>47</sup>Pursuing a Ph.D. at JHU.

- Philip Gianfortoni (2009–10); entered M.S. in Language Technologies at CMU in 2010
- Bill McDowell (2013–4)
- Michael Mordowanec (2014) [246, 394]
- Mourad Heddaya (2020–1); pursuing a Ph.D. at U. Chicago

#### *Mentees at the Allen Institute for AI*

- Gabi Stanovsky (2018–20) young investigator ( $\approx$  postdoc) [89, 229]; lecturer ( $\approx$  assistant professor) at the Hebrew University in Jerusalem
- Suchin Gururangan (2018–20) predoctoral young investigator [77, 84, 92, 97]; Ph.D. student at UW
- Ana Marasović (2019–22) young investigator ( $\approx$  postdoc) [64, 67, 79, 84, 225]; assistant professor at the University of Utah
- Will Merrill (2019–21) predoctoral young investigator; [6, 7, 60, 63, 85]; Ph.D. student at New York U.
- Sarah Wiegrefe (2020) research intern [64]
- Alexander Hoyle (2020) research intern [67]

#### *Career development mentoring*

- Allen School faculty mentor for assistant professors Yulia Tsvetkov (2021–), Tim Althoff (2022–), and Sheng Wang (2022–)
- Keshet Ronen, assistant professor, UW Department of Global Health, NIH K18 awardee (2020–)

### SERVICE

#### *Professional organizations*

- Director (2019–21) and founding board member (2016–), Text as Data Association
- 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; 2018–19)

#### *Journals*

- Editorial board, *Transactions of the Association for Computational Linguistics* (2012–21; paused in 2022)
- Associate editor, *Journal of Artificial Intelligence Research* (2014–19)
- 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*

- General chair, EMNLP 2022
- Action editor, ACL Rolling Review, 2021 (paused in 2022)
- “Computational social science and cultural analytics” senior area chair, ACL 2021
- Area chair, NAACL 2021
- Test of time expert committee, ACL 2020
- “Tagging, chunking, syntax, and parsing” senior area chair, ACL 2019
- Co-organizer (with Jeff Arnold, Emily Gade, and John Wilkerson), [New Directions in Analyzing Text as Data Conference 2018](#)
- 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 [367]
- 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
- “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 and computational linguistics:* ACL 2020, ACL 2018, ACL 2017, ACL 2015, ACL 2014, ACL 2013, ACL 2011, ACL 2010, ACL 2007, COLING-ACL 2006; COLING 2008; CoNLL 2017, CoNLL 2011, CoNLL 2009, CoNLL 2006; EACL 2021, EACL 2017, EACL 2014, EACL 2009; EMNLP 2020, EMNLP 2018, EMNLP 2017, 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 2021, NAACL 2013; SCiL 2018, SCiL 2019
  - Machine learning and artificial intelligence:* ICML 2015, ICML 2013, ICML 2012, ICML 2011, ICML 2008, ICML 2004; IJCAI 2005; NeurIPS 2017, NeurIPS 2010, NeurIPS 2009, NeurIPS 2008, NeurIPS 2007; UAI 2009
  - Other:* WWW 2014, WWW 2011

## *Government agencies (reviewing)*

- National Academies (2017)
- National Science Foundation (2019, 2018, 2017, 2016, 2015, 2012, 2011, 2010, 2009, 2008, 2007, 2006)
- Qatar National Research Foundation (2012, 2011)
- Fundação para a Ciência e a Tecnologia, Portugal (2008)
- Israel Science Foundation (2017)

### *University, School, and Departments*

- CSE co-chair of faculty mentorship and career development (2021–)
- CSE chair of senior faculty mentorship (2020–1)
- CSE graduate admissions committee (2020)
- UW Translation Studies Hub advisory board (2019–)
- CSE undergraduate admissions and scholarships committee (2019)
- CSE executive committee (2017–9)
- CSE faculty recruiting committee (2016–8)
- LTI open house organizer (2014)
- LTI MIIS program advisory board (2013–5)
- 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)

### PROFESSIONAL SOCIETIES

- Member, Association for Computational Linguistics (elected Fellow in 2020)
- Lifetime member, Association for the Advancement of Artificial Intelligence
- Lifetime member, Association for Computing Machinery
- Ally, Black in AI

### PRESENTATIONS

#### *Peer-refereed, non-archival conferences*

- [2020 Annual Meeting of the Society for Political Methodology](#) (given by Sarah Dreier)
- 2019 New Directions in Analyzing Text as Data Conference (given by Lucy Lin) [[356](#)]
- [2018 New Directions in Analyzing Text as Data Conference](#) (given by Sarah Dreier) [[358](#)]
- 2018 International Conference on Computational Social Science (given by Dallas Card) [[111](#)]
- 2018 International Conference on Computational Social Science (given by Chenhao Tan) [[115](#)]
- 2018 International Conference on Computational Social Science (given by Chenhao Tan) [[121](#)]
- [2017 New Directions in Analyzing Text as Data Conference](#) (given by Dallas Card)
- [2016 New Directions in Analyzing Text as Data Conference](#) (given by Dallas Card)
- 2016 International Conference on Computational Social Science (given by Dallas Card) [[123](#)]
- 2016 International Conference on Computational Social Science (given by Yanchuan Sim) [[126](#)]



- 2016 Society of Behavioral Medicine (given by Jason Colditz) [366]
- 2015 New Directions in Analyzing Text as Data Conference (given by Dallas Card)
- 2014 New Directions in Analyzing Text as Data Conference (given by Yanchuan Sim) [145]
- 2014 American Political Science Association (given by Amber Boydston) [368]
- 2014 Annual Meeting of the Comparative Agendas Project (given by Amber Boydston) [368]
- 2013 American Political Science Association (given by Justin Gross) [370]
- 2013 Digital Humanities (given by David Bamman) [372]
- 2013 International Framenet Workshop (given by Nathan Schneider) [280]
- 2012 New Directions in Analyzing Text as Data Conference (given by Tae Yano) [167]
- 2011 New Directions in Text Analysis Conference [177, 180, 289]
- 2011 Linguistic Society of America (given by Jacob Eisenstein) [180]
- 1999 Association for Public Policy Analysis and Management [384]

*Invited talks at conferences, workshops, research meetings, and other events*

- Keynote: EMNLP Workshop on Insights from Negative Results in NLP, 11/2021
- Keynote: Conference on Computer Vision and Pattern Recognition, 6/2021 [66, 68, 73, 77, 80, 353]
- Behavioral Insights from Text Conference, Philadelphia, 1/2020
- Samsung AI Forum, Seoul, 11/2019 [103, 110, 226]
- ACL Workshop on Deep Learning and Formal Languages: Bulding Bridges, Florence, 8/2019 [103, 110, 226]
- Third International Workshop on Symbolic-Neural Learning, Tokyo, 7/2019 [103, 110, 226]
- Microsoft AI for Accessibility Sign Language Recognition & Translation Workshop, Redmond, WA, 2/2019 [11]
- Singapore Symposium on Natural Language Processing, Singapore, 7/2018 [104, 108, 113, 362]
- Midwest Speech and Language Days, South Bend, IN, 5/2018 [104, 108, 110, 113, 362]
- Data Science Summit, University of Washington, 4/2018 [106]
- Keynote: Artificial Intelligence Global Conference, Seoul, South Korea, 12/2017 [117, 120, 362]
- Keynote: Conference on Natural Language Processing and Chinese Computing, Dalian, China, 11/2017 [117, 120, 362]
- EMNLP Workshop on Subword and Character Level Models in NLP, Copenhagen, Denmark, 9/2017 [130, 132, 142, 161, 272]
- PoliInformatics Workshop, Bainbridge Island, WA, 8/2017 [362]
- Keynote: Annual Meeting of the Association for Computational Linguistics, Vancouver, BC, 8/2017 [117, 120, 362]
- Exploring Brain and Computer Sciences: Space, Time, and Collaboration, Champalimaud Foundation/Faber Ventures, Lisbon, Portugal, 7/2017 [panel]
- Representation Learning Workshop, Simons Institute, UC Berkeley, 3/2017 [124, 128, 136]
- Mathematical Analysis of Cultural Expressive Forms: Text Data, Los Angeles, CA, 5/2016 [241]
- Machine Learning Thematic Trimester, Toulouse, France, 9/2015 [152, 153]
- Veteran Affairs State of NLP, Washington, DC, 9/2015 [152, 153]
- Atlanta Computational Social Science Workshop, Atlanta, GA, 11/2014 [145, 157]
- Computational Linguistics in Political Science: What Have You Done for Me Lately?, Mannheim, Germany, 10–11/2014 [interdisciplinary research]



- [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 [32, 167, 173, 180, 189, 195, 248, 281]
- [National Science Foundation CISE CAREER workshop](#), Arlington, TX, 5/2013
- [IARPA FUSE PI meeting](#), College Park, MD, 4/2013 [173]
- [AAAI Spring Symposium on Analyzing Microtext](#), Stanford, CA, 3/2013 [32, 161, 180, 189, 248, 254, 281]
- [World Economic Forum](#), Davos, Switzerland, 1/2013 [167, 180, 189, 195, 256, 281]
- [Insight 3.0: The Web Seen by its Insiders](#) 7/2012, [panel on the technology industry]
- [Workshop on Multilingual Modeling at ACL](#), 7/2012 [169]
- [Lisbon Machine Learning School evening lecture](#), 7/2012 [32, 167, 173]
- [Inducing Linguistic Structure Workshop at NAACL](#), 6/2012 [374]
- [Computer Assisted Reporting Conference](#), 2/2012 [32, 173, 180, 189, 256]
- [Tech@State: Real-Time Awareness](#), U.S. State Department, 2/2012 [32, 173, 180, 189, 195, 256]
- [Text to Text Generation Workshop at ACL](#), 6/2011 [187, 188, 191, 206, 298, 330]
- [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 [189, 195, 197, 256, 257]
- [NAACL-HLT Workshop on #SocialMedia: Computational Linguistics in a World of Social Media](#), 6/2010 [189, 195, 197, 256, 257]
- [NeurIPS 2009 Workshop on Approximate Learning of Large Scale Graphical Models](#), 12/2009 [190, 192, 193, 198]
- [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 [192, 193]
- [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 [206, 207]
- CMU “Andrew’s Leap” computer science high school outreach program talk, 7/2007
- DARPA ISAT study group on Engineering Ensemble Effects, 4/2007 [211]
- [HLT-NAACL Doctoral Consortium](#), 6/2006 [panel on the academic/industrial job market]

*Invited talks at academic, industry, and government research colloquia*

- Megagon Labs, 4/2021 [66, 68, 73, 77, 80, 353]
- University of California San Diego, 4/2021 [66, 68, 73, 77, 80, 353]
- University of Massachusetts Lowell, 4/2021 [66, 68, 73, 77, 80, 353]
- Facebook, 1/2021 [80, 89, 222]
- Amazon, 11/2020 [80, 89, 222]
- Georgetown University, 2/2020 [84, 88]
- Reed College, 11/2019

- Learning Machines Seminar, Cornell Tech, 9/2019 [103, 110, 226]
- Artificial Intelligence Research Center, Tokyo, Japan, 7/2019 [91, 96–98, 101, 226, 228, 229, 231, 234]
- Carnegie Mellon University, 11/2018 [101, 103, 110]
- Allen Institute for Artificial Intelligence, 4/2018 [104, 108, 110, 113, 362]
- Facebook AI Research, 3/2018 [104, 108, 110, 113, 362]
- Politifact, 3/2018
- San Diego State University, 3/2018 [136, 362]
- Stony Brook University, 2/2018 [111, 116, 117, 359]
- Columbia University, 2/2018 [111, 116, 117, 359]
- Samsung, 8/2017 [117, 118, 120]
- Tencent, 8/2017 [117, 120, 362]
- Amazon, 5/2017 [124, 127, 128, 132, 136]
- University of Alberta, 4/2017 [124, 127, 128, 132, 136]
- University of Washington, 4/2017 [124, 127, 128, 132, 136]
- Whitman College, 4/2017 [126, 145, 241]
- University of North Carolina at Chapel Hill, Computer Science Department, 12/2016 [124, 127, 128, 132, 136]
- Bloomberg, 11/2016 [126, 145, 241]
- Cubist Systematic Strategies, 11/2016
- Xerox Research Centre Grenoble, 7/2016 [241, 368]
- Google DeepMind, 7/2016 [241, 368]
- Federal Trade Commission, 7/2016 [241, 368]
- Georgia Institute of Technology, 4/2016 [241, 368]
- New York University, 10/2015 [133, 134, 145]
- IBM T. J. Watson Research Center, 10/2015 [133, 134, 145]
- Microsoft Research, 8/2015 [133, 134, 145]
- University of Toronto, 4/2015 [23, 145, 157]
- University of Texas at Austin, 4/2015 [145, 157]
- Georgetown University, 2/2015 [145, 157]
- Duke University, 2/2015 [145, 157]
- Universität Heidelberg, 11/2014 [23, 145, 157]
- Max-Planck-Institut SWS, 11/2014 [23, 145, 157]
- University of Washington, 4/2014 [153, 157]
- Jump Trading, 3/2014
- Department of Defense, 7/2013 [32, 167, 173]
- eBay Research Labs, 5/2013 [32, 167, 173]
- Cornell University, 3/2013 [32, 167, 173]
- University of Pennsylvania, 12/2012 [32, 167, 173]
- Universitat Politècnica de Catalunya, 7/2012 [32, 167, 173]
- Korea Advanced Institute of Science and Technology, 7/2012 [32, 167, 173, 180]
- Microsoft Research, 5/2012 [32, 167, 173, 180]
- University of Washington, 5/2012 [164, 171, 174]
- University of Maryland, 3/2012 [164, 171, 174]

- Johns Hopkins University, 3/2012 [164, 171, 174]
- Carnegie Mellon University, Language Technologies Institute 25th Anniversary, 10/2011 [32, 173, 177, 180, 189, 195, 197, 256, 257]
- In-Q-Tel, 10/2011 [173, 189, 195, 256, 257]
- IBM T. J. Watson Research Center, 5/2011 [177, 180, 289]
- Toyota Technological Institute at Chicago, 3/2011 [180, 189, 256, 257, 296]
- BlackRock, 3/2011 [180, 189, 256, 257, 296]
- University of North Carolina at Chapel Hill, Political Science Department, 2/2011 [180, 189, 256, 257, 296]
- Carnegie Mellon University, Language Technologies Institute, 9/2010 [189, 195, 197, 256, 257]
- Twitter, 7/2010 [180, 189, 195, 197, 256, 257]
- Facebook, 6/2010 [180, 189, 195, 197, 256, 257]
- HP Labs, Social Computing Group, 6/2010 [180, 189, 195, 197, 256, 257]
- Stanford University, 6/2010 [180, 189, 195, 256, 257]
- University of California Berkeley, 6/2010, [180, 189, 195, 256, 257]
- University of Illinois at Urbana-Champaign, Department of Computer Science, 5/2010 [189, 195, 197, 256, 257]
- University of Texas at Austin, Department of Computer Science, 4/2010 [189, 195, 197, 256, 257]
- Universidade Técnica de Lisboa, Instituto Superior Técnico, 3/2010 [189, 195, 197, 256, 257]
- University of Edinburgh, Institute for Communicating and Collaborative Systems, 3/2010 [189, 195, 197, 256, 257]
- University of Sheffield, Department of Computer Science, 3/2010 [189, 195, 197, 256, 257]
- University of Maryland, Department of Computer Science, 11/2009 [192, 194, 195, 200, 201, 208, 212]
- Princeton University, Department of Computer Science, 10/2009 [194, 200, 208, 212]
- University of Massachusetts at Amherst, Department of Computer Science, 3/2009 [194, 200, 208, 212]
- Universidade Técnica de Lisboa, Instituto Superior Técnico, 7/2008 [206, 302, 304]
- IBM T. J. Watson Research Center, 2/2008 [204, 206, 207]
- Carnegie Mellon University, Language Technologies Institute, 12/2007 [204, 206, 207]
- 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 [212]
- Brown University, Computer Science Department, 10/2005 [209]
- Carnegie Mellon University, Language Technologies Institute, 9/2005 [209]
- Carnegie Mellon University, Center for Automated Learning and Discovery, 9/2005 [212]
- University of Pittsburgh, Computer Science Department, 9/2005 [212]
- Microsoft Research, 8/2004 [213]
- Microsoft Research, 7/2004 [214]
- University of Maryland, Institute for Advanced Computer Study, 4/2002 [215]

## PUBLICATIONS

Citation statistics: h-index 93, i10-index 290, total citations  $\geq 37,000$ , according to [Google Scholar](#) (11/2022).

### *Books, edited volumes, and proceedings*

- [1] Marta R. Costa-jussà, Cristina España-Bonet, Pascale Fung, and Noah A. Smith. [Multilingual and interlingual semantic representations for natural language processing: A brief introduction](#). *Computational Linguistics*, June 2020. Guest editors' introduction to special issue.
- [2] 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.
- [3] Noah A. Smith. [Linguistic Structure Prediction](#). Synthesis Lectures on Human Language Technologies. Morgan and Claypool, May 2011. .

### *Journal articles and book chapters*

- [4] Maarten Sap, Anna Jafarpour, Yejin Choi, Noah A. Smith, James W. Pennebaker, and Eric Horvitz. [Imagined versus remembered stories: Quantifying differences in narrative flow](#). *Proceedings of the National Academy of Sciences*, November 2022.
- [5] Sarah K. Dreier, Emily K. Gade, Dallas Card, and Noah A. Smith. [Patterns of bias: How mainstream media operationalize links between mass shootings and terrorism](#). *Political Communication*, September 2022.
- [6] William Merrill, Ashish Sabharwal, and Noah A. Smith. [Saturated transformers are constant-depth threshold circuits](#). *Transactions of the Association for Computational Linguistics*, 10:843–856, 2022.
- [7] William Merrill, Roy Schwartz, Yoav Goldberg, and Noah A. Smith. [Provable limitations of acquiring meaning from ungrounded form: What will future language models understand?](#) *Transactions of the Association for Computational Linguistics*, 9:1047–1060, 2021.
- [8] Zhaofeng Wu, Hao Peng, and Noah A. Smith. [Infusing finetuning with semantic dependencies](#). *Transactions of the Association for Computational Linguistics*, 9:226–242, 2021.
- [9] Phillip Keung, Julian Salazar, Yichao Lu, and Noah A. Smith. [Unsupervised bitext mining and translation via self-trained contextual embeddings](#). *Transactions of the Association for Computational Linguistics*, 8:828–841, 2020.
- [10] Roy Schwartz, Jesse Dodge, Noah A. Smith, and Oren Etzioni. [Green AI](#). *Communications of the Association for Computing Machinery*, 2020.
- [11] Noah A. Smith. [Contextual word representations: Putting words into computers](#). *Communications of the Association for Computing Machinery*, 63(6), June 2020.
- [12] Dallas Card and Noah A. Smith. [On consequentialism and fairness](#). *Frontiers in Artificial Intelligence*, 3, May 2020.
- [13] Brice D. L. Acree, Yanchuan Sim, Justin H. Gross, Amber E. Boydston, and Noah A. Smith. [Etch-a-sketching: Evaluating the post-primary rhetorical moderation hypothesis](#). *American Politics Research*, 48(1):99–131, January 2020.
- [14] Kelvin Luu, Chenhao Tan, and Noah A. Smith. [Measuring online debaters' persuasive skill from text over time](#). *Transactions of the Association for Computational Linguistics*, 7:537–550, September 2019.
- [15] Shomir Wilson, Florian Schaub, Frederick Liu, Kanthashree Mysore Sathyendra, Daniel Smullen, Sebastian Zimmeck, Rohan Ramanath, Peter Story, Fei Liu, Norman Sadeh, and Noah A. Smith. [Analyzing privacy policies at scale: From crowdsourcing to automated annotations](#). *ACM Transactions on the Web*, 13(1), February 2019.

- [16] Hao Tang, Liang Lu, Lingpeng Kong, Kevin Gimpel, Karen Livescu, Chris Dyer, Noah A. Smith, and Steve Renals. [End-to-end neural segmental models for speech recognition](#). *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1254–1264, December 2017.
- [17] Miguel Ballesteros, Chris Dyer, Yoav Goldberg, and Noah A. Smith. [Greedy transition-based dependency parsing with stack LSTMs](#). *Computational Linguistics*, 43(2), June 2017.
- [18] Jason B. Colditz, Joel Welling, Noah A. Smith, A. Everette James, and Brian A. Primack. [World Vaping Day: Contextualizing vaping culture in online social media using a mixed methods approach](#). *Journal of Mixed Methods Research*, April 2017. doi: 10.1177/1558689817702753.
- [19] Dan Jurafsky, Victor Chahuneau, Bryan R. Routledge, and Noah A. Smith. [Linguistic markers of status in food culture: Bourdieu’s \*Distinction\* in a menu corpus](#). *Cultural Analytics*, October 2016.
- [20] Waleed Ammar, Phoebe 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.
- [21] André F. T. Martins, Mário A. T. Figueiredo, Pedro M. Q. Aguiar, Noah A. Smith, and Eric P. Xing. [AD<sup>3</sup>: Alternating directions dual decomposition for MAP inference in graphical models](#). *Journal of Machine Learning Research*, 16:495–545, March 2015.
- [22] Jacob Eisenstein, Brendan O’Connor, Noah A. Smith, and Eric P. Xing. [Diffusion of language change in social media](#). *PLoS ONE*, November 2014.
- [23] 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.
- [24] Kevin Gimpel and Noah A. Smith. [Phrase dependency machine translation with quasi-synchronous tree-to-tree features](#). *Computational Linguistics*, 40(2), June 2014.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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. .
- [29] Noah A. Smith and André F. T. Martins. [Linguistic structure prediction with the sparseptron](#). *ACM Crossroads*, 19(3):44–48, April 2013.
- [30] Victor Chahuneau, Noah A. Smith, and Chris Dyer. [pycdec: A Python interface to cdec](#). *Prague Bulletin of Mathematical Linguistics*, 98:51–61, October 2012.
- [31] 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.
- [32] David Bamman, Brendan O’Connor, and Noah A. Smith. [Censorship and content deletion in Chinese social media](#). *First Monday*, 17(3), March 2012. .
- [33] 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.
- [34] 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.
- [35] Shay B. Cohen and Noah A. Smith. [Covariance in unsupervised learning of probabilistic grammars](#).



*Journal of Machine Learning Research*, 11:3017–3051, November 2010.

- [36] 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. .
- [37] 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.
- [38] Noah A. Smith and Mark Johnson. [Weighted and probabilistic context-free grammars are equally expressive](#). *Computational Linguistics*, 33(4):477–491, December 2007.
- [39] 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)*

- [40] Suchin Gururangan, Dallas Card, Sarah K. Dreier, Emily K. Gade, Leroy Zhifei Wang, Zeyu Wang, Luke Zettlemoyer, and Noah A. Smith. [Whose language counts as high quality? measuring language ideologies in text data selection](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022**.
- [41] Michael Hassid, Hao Peng, Daniel Rotem, Jungo Kasai, Ivan Montero, Noah A. Smith, and Roy Schwartz. [How much does attention actually attend? questioning the importance of attention in pretrained transformers](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022 Findings**.
- [42] Yushi Hu, Chia-Hsuan Lee, Tianbao Xie, Tao Yu, Mari Ostendorf, and Noah A. Smith. [In-context learning for few-shot dialogue state tracking](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022 Findings**.
- [43] Jungo Kasai, Keisuke Sakaguchi, Ronan Le Bras, Hao Peng, Ximing Lu, Dragomir Radev, Yejin Choi, and Noah A. Smith. [Twist decoding: Diverse generators guide each other](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022**.
- [44] Daniel Khashabi, Gabriel Stanovsky, Jonathan Bragg, Nicholas Lourie, Jungo Kasai, Yejin Choi, Noah A. Smith, and Daniel Weld. [Toward reproducible and standardized human evaluation for text generation](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022**.
- [45] Alisa Liu, Swabha Swayamdipta, Noah A. Smith, and Yejin Choi. <https://arxiv.org/pdf/2201.05955.pdf>. In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022 Findings**.
- [46] Bo-Ru Lu, Yushi Hu, Hao Cheng, Mari Ostendorf, and Noah A. Smith. [Unsupervised learning of hierarchical conversation structure](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022 Findings**.
- [47] Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, Eshaan Pathak, Giannis Karamanolakis, Haizhi Gary Lai, Ishan Virendrabhai Purohit, Ishani Mondal, Jacob William Anderson, Kirby C. Kuznia, Krима Doshi, Kuntal Kumar Pal, Maitreya Patel, Mihir Parmar Mehrad Moradshahi, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Savan Doshi, Shailaja Keyur Sampat, Siddhartha Mishra, Sujan Reddy A, Sumanta Patro, Tanay Dixit, Xudong Shen, Chitta Baral, Yejin Choi, Noah A. Smith, Hannaneh Hajishirzi, and Daniel Khashabi. [Super-NaturalInstructions: Generalization via declarative instructions on 1600+ tasks](#). In *Proceedings of the Conference on Empirical Methods in Natural*

- Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022.**
- [48] Tianbao Xie, Chen Henry Wu, Peng Shi, Ruiqi Zhong, Torsten Scholak, Michihiro Yasunaga, Chien-Sheng Wu, Ming Zhong, Pengcheng Yin, Sida I. Wang, Victor Zhong, Bailin Wang, Chengzu Li, Connor Boyle, Ansong Ni, Ziyu Yao, Dragomir Radev, Caiming Xiong, Lingpeng Kong, Rui Zhang, Noah A. Smith, Luke Zettlemoyer, and Tao Yu. [UnifiedSKG: Unifying and multi-tasking structured knowledge grounding with text-to-text language models](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022.**
  - [49] Suchin Gururangan, Mike Lewis, Ari Holtzman, Noah A. Smith, and Luke Zettlemoyer. [DEMIX layers: Disentangling domains for modular language modeling](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, July 2022. **NAACL 2022.**
  - [50] Jungo Kasai, Keisuke Sakaguchi, Ronan Le Bras, Lavinia Dunagan, Jacob Daniel Morrison, Alexander Fabbri, Yejin Choi, and Noah A. Smith. [Bidimensional leaderboards: Generate and evaluate language hand in hand](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, July 2022. **NAACL 2022.**
  - [51] Jungo Kasai, Keisuke Sakaguchi, Lavinia Dunagan, Jacob Daniel Morrison, Ronan Le Bras, Yejin Choi, and Noah A. Smith. [Transparent human evaluation for image captioning](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, July 2022. **NAACL 2022.**
  - [52] Ximing Lu, Sean Welleck, Peter West, Liwei Jiang, Jungo Kasai, Daniel Khashabi, Ronan Le Bras, Lianhui Qin, Youngjae Yu, Rowan Zellers, Noah A. Smith, and Yejin Choi. [Neurologic A\\*esque decoding: Constrained text generation with lookahead heuristics](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, July 2022. **NAACL 2022.**
- Best new method paper award.**
- [53] Kelvin Luu, Daniel Khashabi, Suchin Gururangan, Karishma Mandyam, and Noah A. Smith. [Time waits for no one! analysis and challenges of temporal misalignment](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, July 2022. **NAACL 2022.**
  - [54] Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi, and Noah A. Smith. [Annotators with attitudes: How annotator beliefs and identities bias toxic language detection](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Seattle, WA, July 2022. **NAACL 2022.**
  - [55] Jesse Dodge, Taylor Prewitt, Remi Tachet des Combes, Erika Odmark, Roy Schwartz, Emma Strubell, Alexandra Sasha Luccioni, Noah A. Smith, Nicole DeCario, and Will Buchanan. [Measuring machine learning software carbon intensity in cloud instances](#). In *Conference on Fairness, Accountability, and Transparency*, Seoul, June 2022. **FAccT 2022.**
  - [56] Tal August, Katharina Reinecke, and Noah A. Smith. [Generating scientific definitions with controllable complexity](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Dublin, May 2022. **ACL 2022.**
  - [57] Yao Dou, Maxwell Forbes, Rik Koncel-Kedziorski, Noah A. Smith, and Yejin Choi. [Is GPT-3 text indistinguishable from human text? SCARECROW: A framework for scrutinizing machine text](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Dublin, May 2022. **ACL 2022.**
  - [58] Hao Peng, Jungo Kasai, Nikolaos Pappas, Dani Yogatama, Zhaofeng Wu, Lingpeng Kong, Roy Schwartz, and Noah A. Smith. [ABC: Attention with bounded-memory control](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Dublin, May 2022. **ACL 2022.**



- [59] Ofir Press, Noah A. Smith, and Mike Lewis. [Train short, test long: Attention with linear biases enables input length extrapolation](#). In *Proceedings of the International Conference on Learning Representations*, virtual conference, April 2022. **ICLR 2022**.
- [60] Matt Gardner, William Merrill, Jesse Dodge, Matthew Peters, Alexis Ross, Sameer Singh, and Noah A. Smith. [Competency problems: On finding and removing artifacts in language data](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November 2021. **EMNLP 2021**.
- [61] Jungo Kasai, Hao Peng, Yizhe Zhang, Dani Yogatama, Gabriel Ilharco, Nikolaos Pappas, Yi Mao, Weizhu Chen, and Noah A. Smith. [Finetuning pretrained transformers into RNNs](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November 2021. **EMNLP 2021**.
- [62] Leo Zeyu Liu, Yizhong Wang, Jungo Kasai, Hannaneh Hajishirzi, and Noah A. Smith. [Probing across time: What does RoBERTa know and when?](#) In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November 2021. **EMNLP 2021 Findings**.
- [63] William Merrill, Vivek Ramanujan, Yoav Goldberg, Roy Schwartz, and Noah A. Smith. [Effects of parameter norm growth during transformer training: Inductive bias from gradient descent](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November 2021. **EMNLP 2021**.
- [64] Sarah Wiegreffe, Ana Marasović, and Noah A. Smith. [Measuring association between labels and free-text rationales](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November 2021. **EMNLP 2021**.
- [65] Rahul Nadkarni, David Wadden, Iz Beltagy, Noah A. Smith, Hannaneh Hajishirzi, and Tom Hope. [Scientific language models for biomedical knowledge base completion: An empirical study](#). In *Proceedings of the Conference on Automated Knowledge Base Construction*, Irvine, CA, October 2021. **AKBC 2021**.
- [66] Elizabeth Clark, Tal August, Sofia Serrano, Nikita Haduong, Suchin Gururangan, and Noah A. Smith. [All that’s ‘human’ is not gold: Evaluating human evaluation of generated text](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, August 2021. **ACL 2021**.  
**Outstanding paper award.**
- [67] Alexander Miserlis Hoyle, Ana Marasović, and Noah A. Smith. [Promoting graph awareness in linearized graph-to-text generation](#). In *Findings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, August 2021. **ACL 2021 Findings**.
- [68] Alisa Liu, Maarten Sap, Ximing Lu, Swabha Swayamdipta, Chandra Bhagavatula, Noah A. Smith, and Yejin Choi. [DExperts: On-the-fly controlled text generation with experts and anti-experts](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, August 2021. **ACL 2021**.
- [69] Kelvin Luu, Xinyi Wu, Rik Koncel-Kedziorski, Kyle Lo, Isabel Cachola, and Noah A. Smith. [Explaining relationships between scientific documents](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, August 2021. **ACL 2021**.
- [70] Ofir Press, Noah A. Smith, and Mike Lewis. [Shortformer: Better language modeling using shorter inputs](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, August 2021. **ACL 2021**.
- [71] Pradeep Dasigi, Kyle Lo, Iz Beltagy, Arman Cohan, Noah A. Smith, and Matt Gardner. [A dataset of information-seeking questions and answers anchored in research papers](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, virtual

- conference, June 2021. **NAACL 2021.**
- [72] Jungo Kasai, Nikolaos Pappas, Hao Peng, James Cross, and Noah A. Smith. [Deep encoder, shallow decoder: Reevaluating non-autoregressive machine translation](#). In *Proceedings of the International Conference on Learning Representations*, virtual conference, May 2021. **ICLR 2021.**
  - [73] Hao Peng, Nikolaos Pappas, Dani Yogatama, Roy Schwartz, Noah A. Smith, and Lingpeng Kong. [Random feature attention](#). In *Proceedings of the International Conference on Learning Representations*, virtual conference, May 2021. **ICLR 2021.**
  - [74] Xuhui Zhou, Maarten Sap, Swabha Swayamdipta, Yejin Choi, and Noah A. Smith. [Challenges in algorithmic debiasing for toxic language detection](#). In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, virtual conference, April 2021. **EACL 2021.**
  - [75] Tal August, Lauren Kim, Katharina Reinecke, and Noah A. Smith. [Writing strategies for science communication: Data and computational analysis](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020.**
  - [76] Matt Gardner, Yoav Artzi, Victoria Basmova, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, Nitish Gupta, Hannaneh Hajishirzi, Gabriel Ilharco, Daniel Khoshnab, Kevin Lin, Jiangming Liu, Nelson F. Liu, Phoebe Mulcaire, Qiang Ning, Sameer Singh, Noah A. Smith, Sanjay Subramanian, Reut Tsarfay, Eric Wallace, Ally Zhang, and Ben Zhou. [Evaluating NLP models via contrast sets](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020 Findings.**
  - [77] Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. [RealToxicityPrompts: Evaluating neural toxic degeneration in language models](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020 Findings.**
  - [78] Florian Mai, Nikolaos Pappas, Ivan Montero, Noah A. Smith, and James Henderson. [Plug and play autoencoders for conditional text generation](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020.**
  - [79] Ana Marasović, Chandra Bhagavatula, Jae Sung Park, Ronan Le Bras, Noah A. Smith, and Yejin Choi. [Natural language rationales with full-stack visual reasoning: From pixels to semantic frames to commonsense graphs](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020 Findings.**
  - [80] Nikolaos Pappas, Phoebe Mulcaire, and Noah A. Smith. [Grounded compositional outputs for adaptive language modeling](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020.**
  - [81] Swabha Swayamdipta, Roy Schwartz, Nicholas Lourie, Yizhong Wang, Hannaneh Hajishirzi, Noah A. Smith, and Yejin Choi. [Dataset cartography: Mapping and diagnosing datasets with training dynamics](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020.**
  - [82] Rachel Rudinger, Vered Shwartz, Jena D. Hwang, Chandra Bhagavatula, Maxwell Forbes, Ronan Le Bras, Noah A. Smith, and Yejin Choi. [Thinking like a skeptic: Defeasible inference in natural language](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020 Findings.**
  - [83] Xuhui Zhou, Nikolaos Pappas, and Noah A. Smith. [Multilevel text alignment with cross-document attention](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020.**
  - [84] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and

Noah A. Smith. [Don't stop pretraining: Adapt language models to domains and tasks](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.

**Honorable mention for best paper award.**

- [85] William Merrill, Gail Weiss, Yoav Goldberg, Roy Schwartz, Noah A. Smith, and Eran Yahav. [A formal hierarchy of RNN architectures](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.
- [86] Hao Peng, Roy Schwartz, Dianqi Li, and Noah A. Smith. [A mixture of  \$h - 1\$  heads is better than  \$h\$  heads](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.
- [87] Ofir Press, Noah A. Smith, and Omer Levy. [Improving transformer models by reordering their sublayers](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.
- [88] Maarten Sap, Saadia Gabriel, Lianhui Qin, Dan Jurafsky, Noah A. Smith, and Yejin Choi. [Social bias frames: Reasoning about social and power implications of language](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.
- [89] Roy Schwartz, Gabriel Stanovsky, Swabha Swayamdipta, Jesse Dodge, and Noah A. Smith. [The right tool for the job: Matching model and instance complexities](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.
- [90] Tal August, Dallas Card, Gary Hsieh, Noah A. Smith, and Katharina Reinecke. [Explain like I am a scientist: The linguistic barriers of entry to r/science](#). In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, April 2020. **CHI 2020**.
- [91] Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. [Low-resource parsing with crosslingual contextualized representations](#). In *Proceedings of the Conference on Computational Natural Language Learning*, Hong Kong, November 2019. **CoNLL 2019**.
- [92] Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, and Noah A. Smith. [Show your work: Improved reporting of experimental results](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [93] Sachin Kumar, Shuly Wintner, Noah A. Smith, and Yulia Tsvetkov. [Topics to avoid: Demoting latent confounds in text classification](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [94] XiuJun Li, Chunyuan Li, Qiaolin Xia, Yonatan Bisk, Asli Celikyilmaz, Jianfeng Gao, Noah A. Smith, and Yejin Choi. [Robust navigation with language pretraining and stochastic sampling](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [95] Matthew E. Peters, Mark Neumann, Robert L. Logan IV, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. [Knowledge enhanced contextual word representations](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [96] Elizabeth Clark, Asli Celikyilmaz, and Noah A. Smith. [Sentence mover's similarity: Automatic evaluation for multi-sentence texts](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. **ACL 2019**.
- [97] Suchin Gururangan, Tam Dang, Dallas Card, and Noah A. Smith. [Variational pretraining for semi-supervised text classification](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. **ACL 2019**.
- [98] Sofia Serrano and Noah A. Smith. [Is attention interpretable?](#) In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. **ACL 2019**.

- [99] Nelson F. Liu, Matt Gardner, Yonatan Belinkov, Matthew E. Peters, and Noah A. Smith. [Linguistic knowledge and transferability of contextual representations](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Minneapolis, MN, June 2019. **NAACL 2019.**
- [100] Maarten Sap, Ronan LeBras, Emily Allaway, Chandra Bhagavatula, Nicholas Lourie, Hannah Rashkin, Brendan Roof, Noah A. Smith, and Yejin Choi. [ATOMIC: An atlas of machine common-sense for if-then reasoning](#). In *Proceedings of AAAI Conference on Artificial Intelligence*, Honolulu, HI, January/February 2019. **AAAI 2019.**
- [101] Dallas Card, Michael Zhang, and Noah A. Smith. [Deep weighted averaging classifiers](#). In *Proceedings of the ACM Fairness, Accountability, and Transparency Conference*, New York, NY, January 2019. **FAT\* 2019.**
- [102] Tal August, Nigini Oliveira, Chenhao Tan, Noah A. Smith, and Katharina Reinecke. [Framing effect: The choice of slogans used to advertise online experiments can boost recruitment and lead to sample biases](#). In *Proceedings of the Conference on Computer-Supported Cooperative Work and Social Computing*, Jersey City, NJ, November 2018. **CSCW 2018.**
- [103] Hao Peng, Roy Schwartz, Sam Thomson, and Noah A. Smith. [Rational recurrences](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, November 2018. **EMNLP 2018.**
- [104] Swabha Swayamdipta, Sam Thomson, Kenton Lee, Luke Zettlemoyer, Chris Dyer, and Noah A. Smith. [Syntactic scaffolds for semantic structures](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, November 2018. **EMNLP 2018.**
- [105] Jiateng Xie, Zhilin Yang, Graham Neubig, Noah A. Smith, and Jaime Carbonell. [Neural cross-lingual named entity recognition with minimal resources](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium, November 2018. **EMNLP 2018.**
- [106] Lucy H. Lin, Scott B. Miles, and Noah A. Smith. [Natural language processing for analyzing disaster recovery trends expressed in large text corpora](#). In *IEEE Global Humanitarian Technology Conference*, San Jose, California, October 2018. **GHTC 2018.**
- [107] Dallas Card, Chenhao Tan, and Noah A. Smith. [Neural models for documents with metadata](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July 2018. **ACL 2018.**
- [108] Hao Peng, Sam Thomson, and Noah A. Smith. [Backpropagating through structured argmax using a SPIGOT](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July 2018. **ACL 2018. Honorable mention for best paper award.**
- [109] Hannah Rashkin, Maarten Sap, Emily Allaway, Noah A. Smith, and Yejin Choi. [Event2Mind: Commonsense inference on events, intents, and reactions](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July 2018. **ACL 2018.**
- [110] Roy Schwartz, Sam Thomson, and Noah A. Smith. [Bridging CNNs, RNNs, and weighted finite-state machines](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July 2018. **ACL 2018.**
- [111] Dallas Card and Noah A. Smith. [The importance of calibration for estimating proportions from annotations](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, New Orleans, LA, June 2018. **NAACL 2018.**
- [112] Elizabeth Clark, Yangfeng Ji, and Noah A. Smith. [Neural text generation in stories using entity representations as context](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, New Orleans, LA, June 2018. **NAACL 2018. Outstanding paper award.**



- [113] Hao Peng, Sam Thomson, Swabha Swayamdipta, and Noah A. Smith. [Learning joint semantic parsers from disjoint data](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, New Orleans, LA, June 2018. **NAACL 2018**.
- [114] Yijia Liu, Yi Zhu, Wanxiang Che, Bing Qin, Nathan Schneider, and Noah A. Smith. [Parsing tweets into universal dependencies](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, New Orleans, LA, June 2018. **NAACL 2018**.
- [115] Chenhao Tan, Hao Peng, and Noah A. Smith. [“You are no Jack Kennedy”: On media highlights of presidential debates](#). In *Proceedings of the International World Wide Web Conference*, Lyon, France, April 2018. **WWW 2018**.
- [116] Elizabeth Clark, Anne Spencer Ross, Chenhao Tan, Yangfeng Ji, and Noah A. Smith. [Creative writing with a machine in the loop: Case studies on slogans and stories](#). In *Proceedings of the Conference on Intelligent User Interfaces*, Tokyo, Japan, March 2018. **IUI 2018**.
- [117] Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, and Noah A. Smith. [Dynamic entity representations in neural language models](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Copenhagen, Denmark, September 2017. **EMNLP 2017**.
- [118] Roy Schwartz, Maarten Sap, Ioannis Konstas, Leila Zilles, Yejin Choi, and Noah A. Smith. [The effect of different writing tasks on linguistic style: A case study of the ROC story cloze task](#). In *Proceedings of the Conference on Computational Natural Language Learning*, Vancouver, BC, August 2017. **CoNLL 2017**.
- [119] Yangfeng Ji and Noah A. Smith. [Neural discourse structure for text categorization](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Vancouver, BC, July/August 2017. **ACL 2017**.
- [120] Hao Peng, Sam Thomson, and Noah A. Smith. [Deep multitask learning for semantic dependency parsing](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Vancouver, BC, July/August 2017. **ACL 2017**.
- [121] Chenhao Tan, Dallas Card, and Noah A. Smith. [Friendships, rivalries, and trysts: Characterizing relations between ideas in texts](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Vancouver, BC, July/August 2017. **ACL 2017**.
- [122] Adhiguna Kuncoro, Miguel Ballesteros, Lingpeng Kong, Chris Dyer, Graham Neubig, and Noah A. Smith. [What do recurrent neural network grammars learn about syntax?](#) In *Proceedings of the Conference of the European Chapter of the Association for Computational Linguistics*, Valencia, Spain, April 2017. **EACL 2017**.  
**Outstanding paper award.**
- [123] Dallas Card, Justin H. Gross, Amber E. Boydstun, 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.
- [124] 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**.
- [125] 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**.
- [126] 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.

- [127] 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**.
- [128] 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**.
- [129] 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**.
- [130] 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**.
- [131] 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**.
- [132] 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**.
- [133] 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**.
- [134] 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**.
- [135] 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**.
- [136] 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**.
- [137] Manaal Faruqui, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah A. Smith. [Sparse over-complete word vector representations](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Beijing, China, July 2015. **ACL 2015**.
- [138] 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**.
- [139] 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**.
- [140] 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**.
- [141] 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**.
- [142] 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**.
  - [143] 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**.
  - [144] 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**.
  - [145] 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**.
  - [146] 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).**
  - [147] 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**.
  - [148] 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**.
  - [149] 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**.
  - [150] 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**.
  - [151] 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.**
  - [152] 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**.
  - [153] 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**.
  - [154] 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**.
  - [155] 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**.



- [156] 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**.
- [157] 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**.
- [158] 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**.
- [159] 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**.
- [160] 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**.
- [161] 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*, Atlanta, GA, June 2013. **NAACL 2013**.
- [162] 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**.
- [163] 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**.
- [164] 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**.
- [165] 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**.
- [166] 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**.
- [167] 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**.
- [168] 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**.
- [169] 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**.
- [170] 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.**
- [171] 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.**
- [172] 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.**
- [173] 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.**
- [174] 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.**
- [175] 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.**
- [176] 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.**
- [177] 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.**
- [178] 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.**
- [179] 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.**
- [180] 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.**
- [181] 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.**
- [182] 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.**
- [183] 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.**
- [184] 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.**
- [185] 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.**
- [186] 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.**
  - [187] 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.**
  - [188] 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.**
  - [189] 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.**
  - [190] 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.**
  - [191] Dipanjan Das and Noah A. Smith. [Paraphrase identification as probabilistic quasi-synchronous recognition](#). 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.**
  - [192] 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.**
  - [193] 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.**
  - [194] 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.**
  - [195] 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.**
  - [196] 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.**
  - [197] 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.**

- [198] 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**.
- [199] 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.**
- [200] 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**.
- [201] 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**.
- [202] 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**.
- [203] 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**.
- [204] Shay B. Cohen and Noah A. Smith. [Joint morphological and syntactic disambiguation](#). In *Proceedings 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**.
- [205] 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**.
- [206] 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.**
- [207] 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**.
- [208] 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**.
- [209] 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**.
- [210] 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**.



- [211] 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.**
- [212] 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.**
- [213] 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.**
- [214] 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.**
- [215] 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)*

- [216] Zhaofeng Wu, Hao Peng, Nikolaos Pappas, and Noah A. Smith. [Modeling context with linear attention for scalable document-level translation](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, December 2022. **EMNLP 2022 Findings.**
- [217] Daniel Cheng, Kyle Yan, Phillip Keung, and Noah A. Smith. [The engage corpus: A social media dataset for text-based recommender systems](#). In *Proceedings of the Language Resources and Evaluation Conference*, Marseille, France, June 2022. **LREC 2022.**
- [218] Daniel Edmiston, Phillip Keung, and Noah A. Smith. [Domain mismatch doesn’t always prevent cross-lingual transfer learning](#). In *Proceedings of the Language Resources and Evaluation Conference*, Marseille, France, June 2022. **LREC 2022.**
- [219] Jesse Dodge, Suchin Gururangan, Dallas Card, Roy Schwartz, and Noah A. Smith. [Expected validation performance and estimation of a random variable’s maximum](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, Punta Cana, Dominican Republic, November 2021. **EMNLP 2021 Findings.**
- [220] Ivan Montero, Nikolaos Pappas, and Noah A. Smith. [Sentence bottleneck autoencoders from transformer language models](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, November 2021. **EMNLP 2021.**
- [221] Elizabeth Clark and Noah A. Smith. [Choose your own adventure: Paired suggestions in collaborative writing for evaluating story generation models](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, virtual conference, June 2021. **NAACL 2021.**
- [222] Ethan C. Chau, Lucy H. Lin, and Noah A. Smith. [Parsing with multilingual BERT, a small treebank, and a small corpus](#). In *Findings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020 Findings.**
- [223] Phillip Keung, Yichao Lu, György Szarvas, and Noah A. Smith. [The multilingual Amazon reviews corpus](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, virtual conference, November 2020. **EMNLP 2020.**
- [224] Maarten Sap, Eric Horvitz, Yejin Choi, Noah A. Smith, and James Pennebaker. [Recollection versus](#)

- imagination: Exploring human memory and cognition via large language models. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, virtual conference, July 2020. **ACL 2020**.
- [225] Pradeep Dasigi, Nelson F. Liu, Ana Marasović, Noah A. Smith, and Matt Gardner. [Quoref: A reading comprehension dataset with questions requiring coreferential reasoning](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [226] Jesse Dodge, Roy Schwartz, Hao Peng, and Noah A. Smith. [RNN architecture learning with sparse regularization](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [227] Hao Peng, Roy Schwartz, and Noah A. Smith. [PaLM: A hybrid parser and language model](#). In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Hong Kong, November 2019. **EMNLP 2019**.
- [228] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A. Smith. [The risk of racial bias in hate speech detection](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. **ACL 2019**.  
**Nominated for best short paper award.**
- [229] Gabriel Stanovsky, Noah A. Smith, and Luke Zettlemoyer. [Evaluating gender bias in machine translation](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 2019. **ACL 2019**.  
**Nominated for best short paper award.**
- [230] Nelson F. Liu, Roy Schwartz, and Noah A. Smith. [Inoculation by fine-tuning: A method for analyzing challenge datasets](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Minneapolis, MN, June 2019. **NAACL 2019**.
- [231] Phoebe Mulcaire, Jungo Kasai, and Noah A. Smith. [Polyglot contextual representations improve crosslingual transfer](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, Minneapolis, MN, June 2019. **NAACL 2019**.
- [232] Phoebe Mulcaire, Swabha Swayamdipta, and Noah A. Smith. [Polyglot semantic role labeling](#). In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, Melbourne, Australia, July 2018. **ACL 2018**.
- [233] Hao Fang, Hao Cheng, Maarten Sap, Elizabeth Clark, Ari Holtzman, Yejin Choi, Noah A. Smith, and Mari Ostendorf. [Sounding Board: A user-centric and content-driven social chatbot](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (Demonstration Papers)*, New Orleans, LA, June 2018. **NAACL 2018**.
- [234] Suchin Gururangan, Swabha Swayamdipta, Omer Levy, Roy Schwartz, Samuel R. Bowman, and Noah A. Smith. [Annotation artifacts in natural language inference data](#). In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics*, New Orleans, LA, June 2018. **NAACL 2018**.
- [235] Liang Lu, Lingpeng Kong, Chris Dyer, and Noah A. Smith. [Multitask learning with CTC and segmental CRF for speech recognition](#). In *Proceedings of InterSpeech*, Stockholm, Sweden, August 2017. **InterSpeech 2017**.
- [236] 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**.
- [237] 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**.

- [238] 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**.
- [239] 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**.
- [240] 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**.
- [241] 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**.
- [242] 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**.
- [243] 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**.
- [244] 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**.
- [245] 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**.
- [246] 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**.
- [247] 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**.
- [248] 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**.
- [249] 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**.
- [250] 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**.
- [251] 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**.
- [252] 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**.



- [253] 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**.
- [254] 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**.
- [255] 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**.
- [256] 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**.
- [257] 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**.
- [258] 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**.
- [259] 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**.
- [260] 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**.
- [261] 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**.
- [262] 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**.
- [263] 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**.
- [264] 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*

- [265] Ethan C. Chau and Noah A. Smith. [Specializing multilingual language models: An empirical study](#). In *Proceedings of the EMNLP Workshop on Multilingual Representation Learning*, Punta Cana, Dominican Republic, November 2021. **MRL 2021.**  
**Honorable mention for best paper award.**
- [266] Tal August, Maarten Sap, Elizabeth Clark, Katharina Reinecke, and Noah A. Smith. [Exploring the effect of author and reader identity in online story writing: the storiesinthewild corpus](#). In *ACL Workshop on Narrative Understanding, Storylines, and Events*, July 2020. **NUSE 2020.**
- [267] Matthew Peters, Sebastian Ruder, and Noah A. Smith. [To tune or not to tune? adapting pretrained representations to diverse tasks](#). In *Proceedings of the ACL Workshop on Representation Learning for Natural Language Processing*, Florence, Italy, August 2019. **Repl4NLP 2019.**
- [268] Nelson F. Liu, Omer Levy, Roy Schwartz, Chenhao Tan, and Noah A. Smith. [LSTMs exploit linguistic attributes of data](#). In *Proceedings of the ACL Workshop on Representation Learning for Natural Language Processing*, Melbourne, Australia, July 2018. **RepL4NLP 2018.**  
**Best paper award.**
- [269] Nelson F. Liu, Gina-Anne Levow, and Noah A. Smith. [Discovering phonesthemes with sparse regularization](#). In *Proceedings of the NAACL Workshop on Subword and Character-Level Models in Natural Language Processing*, New Orleans, LA, June 2018.
- [270] Roy Schwartz, Maarten Sap, Ioannis Konstas, Li Zilles, Yejin Choi, and Noah A. Smith. [Story cloze task: UW NLP system](#). In *Proceedings of the Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, pages 52–55, Valencia, Spain, April 2017. **LSDSem 2017.**
- [271] Aaron Jaech, Phoebe 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.**
- [272] Aaron Jaech, Phoebe 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.**
- [273] Jeffrey Flanigan, Chris Dyer, Noah A. Smith, and Jaime Carbonell. [CMU at SemEval-2016 task 8: Graph-based AMR parsing with infinite ramp loss](#). In *Proceedings of the NAACL Workshop on Semantic Evaluations*, San Diego, CA, June 2016. **SemEval 2016.**
- [274] Mohammad Javad Hosseini, Noah A. Smith, and Su-In Lee. [UW-CSE: Detecting multiword expressions and supersenses using double-chained conditional random fields](#). In *Proceedings of the NAACL Workshop on Semantic Evaluations*, San Diego, CA, June 2016. **SemEval 2016.**
- [275] 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.
- [276] Rohan Ramanath, Florian Schaub, Shomir Wilson, Fei Liu, Norman Sadeh, and Noah A. Smith. [Identifying relevant text fragments to help crowdsource privacy policy annotations](#). In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, Pittsburgh, PA, November 2014. **HCOMP 2014.**
- [277] 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.**
- [278] 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.
- [279] 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**.
- [280] Nathan Schneider, Chris Dyer, and Noah A. Smith. Exploiting and expanding corpus resources for frame-semantic parsing. April 2013. International FrameNet Workshop.
- [281] 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.
- [282] 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.
- [283] 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.
- [284] 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.
- [285] 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**.
- [286] 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**.
- [287] 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**.
- [288] 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**.
- [289] 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**.
- [290] 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**.
- [291] 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.
- [292] 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.
- [293] 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**.
- [294] 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.
- [295] 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.
- [296] 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.
  - [297] 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.
  - [298] 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.
  - [299] 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.
  - [300] 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.
  - [301] 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.
  - [302] 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.**
  - [303] 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*

- [304] 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.
- [305] 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.
- [306] 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.
- [307] 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*

- [308] Hao Peng. *Towards Efficient and Generalizable Natural Language Processing*. Ph.D. thesis, University of Washington, Seattle, WA, May 2022.
- [309] Phoebe Mulcaire. *Crosslingual Sharing for Low-Resource Natural Language Processing*. Ph.D. thesis, University of Washington, Seattle, WA, May 2022.
- [310] Lucy H. Lin. *Semantic Comparisons for Natural Language Processing Applications*. Ph.D. thesis, University of Washington, Seattle, WA, December 2021.
- [311] Maarten Sap. *Positive AI with Social Commonsense Models*. Ph.D. thesis, University of Washington, Seattle, WA, July 2021.
- Awarded the William Chan Memorial Dissertation Award.**
- [312] Elizabeth Clark. *Text Generation and Evaluation for Human-Machine Collaborative Writing*. Ph.D. thesis, University of Washington, Seattle, WA, June 2021.
- [313] Jesse Dodge. *Towards Efficient and Reproducible Natural Language Processing*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, May 2020.
- [314] Dallas Card. *Accelerating Text-as-Data Research in Computational Social Science*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, August 2019.
- [315] Swabha Swayamdipta. *Syntactic Inductive Biases for Natural Language Processing*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, May 2019.
- [316] Sam Thomson. *Encoding and Decoding Graph Representations of Natural Language*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, February 2019.
- [317] Jeffrey Flanigan. *Parsing and Generation for the Abstract Meaning Representation*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, November 2018.
- [318] Lingpeng Kong. *Neural Representation Learning in Linguistic Structured Prediction*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, September 2017.
- [319] Yanchuan Sim. *Text as Strategic Choice*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, June 2016.
- [320] Waleed Ammar. *Towards a Universal Analyzer of Natural Languages*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, June 2016.
- [321] 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.**
- [322] Dani Yogatama. *Sparse Models of Natural Language Text*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, May 2015.
- [323] Brendan O'Connor. *Statistical Text Analysis for Social Science*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, August 2014.
- [324] Nathan Schneider. *Lexical Semantic Analysis in Natural Language Text*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, June 2014.
- [325] Tae Yano. *Text as Actuator: Text-Driven Response Modeling and Prediction in Politics*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, July 2013.
- [326] Kevin Gimpel. *Discriminative Feature-Rich Modeling for Syntax-Based Machine Translation*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, August 2012.
- [327] 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.**
- [328] Dipanjan Das. *Semi-Supervised and Latent-Variable Models of Natural Language Semantics*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, April 2012.



- [329] Shay B. Cohen. *Computational Learning of Probabilistic Grammars in the Unsupervised Setting*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, September 2011.
- [330] Michael Heilman. *Automatic Factual Question Generation from Text*. Ph.D. thesis, Carnegie Mellon University, Pittsburgh, PA, April 2011.

*Supervised undergraduate and masters theses*

- [331] Tobias Rohde. Integrating inductive biases into transformers. Masters thesis, University of Washington, June 2022.
- [332] Ivan Montero. Sentence bottleneck autoencoders from transformer language models. Masters thesis, University of Washington, June 2022.
- [333] Zhilin Wang. Extracting and inferring personal attributes from dialogue. Masters thesis, University of Washington, August 2021.
- [334] Ethan Chau. Specializing multilingual language models: An empirical study. Masters thesis, University of Washington, May 2021.
- [335] Zeyu Blarry Wang. Difficulties in detecting depression symptoms from clinical SMS in Kenya with NLP. Masters thesis, University of Washington, May 2021.
- [336] Zhaofeng Wu. Infusing finetuning with semantic dependencies. Masters thesis, University of Washington, May 2021.
- [337] Xuhui Zhou. Challenges in automated debiasing for toxic language detection. Masters thesis, University of Washington, May 2020.
- [338] Ethan C. Chau. Towards resource-efficient contextual word representations for parsing, June 2020. Computer Science honors thesis, Paul G. Allen School of Computer Science & Engineering, University of Washington.
- [339] Sam Gehman. Investigating language model training data. Masters thesis, University of Washington, June 2020.
- [340] Karishma Mandyam. Semi-supervised learning for verbal autopsy. Masters thesis, University of Washington, June 2020.
- [341] Tam Dang. Modern variational methods for semi-supervised text classification. Masters thesis, University of Washington, Seattle, WA, June 2019.
- [342] Ron Fan. Topic modeling with semantic frames. Masters thesis, University of Washington, Seattle, WA, June 2019.
- [343] Nelson F. Liu. Implicit linguistic knowledge of neural natural language processing models, June 2019. Computer Science honors thesis, Paul G. Allen School of Computer Science & Engineering, University of Washington.
- [344] Karishma Mandyam. Linguistic preprocessing for attribute transfer, June 2019. Computer Science honors thesis, Paul G. Allen School of Computer Science & Engineering, University of Washington.
- [345] Aishwarya Nirmal. Inferring gender from context: Comparing humans and pretrained models. Masters thesis, University of Washington, Seattle, WA, June 2019.
- [346] Deric Pang. Improving natural language inference with syntactic word representations. Masters thesis, University of Washington, Seattle, WA, June 2019.
- [347] Suchin Gururangan. Polyglot text classification with neural document models. Masters thesis, University of Washington, Seattle, WA, August 2018.
- [348] Yi Zhu. Dependency parsing for tweets. Masters thesis, University of Washington, Seattle, WA, August 2017.
- [349] Desai Chen. Unsupervised bilingual POS tagging with Markov random fields, May 2011. Computer Science honors thesis, School of Computer Science, Carnegie Mellon University.

- [350] Rishav Bhowmick. [Rich entity type recognition in text](#), May 2010. Computer Science honors thesis, Department of Computer Science, Carnegie Mellon University–Qatar.
- [351] Neel Shah. [Predicting risk from financial reports with supervised topic models](#), May 2010. Computer Science honors thesis, School of Computer Science, Carnegie Mellon University.
- [352] 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*

- [353] Daniel Khashabi, Gabriel Stanovsky, Jonathan Bragg, Nicholas Lourie, Jungo Kasai, Yejin Choi, Noah A. Smith, and Daniel S. Weld. [GENIE: A leaderboard for human-in-the-loop evaluation of text generation](#), January 2021.
- [354] Qiaolin Xia, Xiujuan Li, Chunyuan Li, Yonatan Bisk, Zhifang Sui, Jianfeng Gao, Yejin Choi, and Noah A. Smith. [Multi-view learning for vision-and-language navigation](#), March 2020.
- [355] Jesse Dodge, Gabriel Ilharco, Roy Schwartz, Ali Farhadi, Hannaneh Hajishirzi, and Noah A. Smith. [Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping](#), February 2020.
- [356] Lucy H. Lin and Noah A. Smith. [Situating sentence embedders with nearest neighbor overlap](#), September 2019.
- [357] Swabha Swayamdipta, Matthew Peters, Brendan Roof, Chris Dyer, and Noah A. Smith. [Shallow syntax in deep water](#), August 2019.
- [358] Sarah Dreier, Lucy H. Lin, Sofia Serrano, Emily Gade, and Noah A. Smith. Religiosity and public policy in congress: A text analysis of U.S. federal legislators’ religious rhetoric and policy attitudes. Technical report, September 2018.
- [359] Lucy H. Lin, Scott B. Miles, and Noah A. Smith. [Semantic matching against a corpus: New methods and applications](#), August 2018.
- [360] Dallas Card, Lucy H. Lin, and Noah A. Smith. [Politifact language audit](#), March 2018.
- [361] Ofir Press and Noah A. Smith. [You may not need attention](#), October 2018.
- [362] Swabha Swayamdipta, Sam Thomson, Chris Dyer, and Noah A. Smith. [Frame-semantic parsing with softmax-margin segmental RNNs and a syntactic scaffold](#), June 2017.
- [363] Jesse Dodge, Kevin Jamieson, and Noah A. Smith. [Open loop hyperparameter optimization and determinantal point processes](#), June 2017.
- [364] Waleed Ammar, Phoebe Mulcaire, Yulia Tsvetkov, Guillaume Lample, Chris Dyer, and Noah A. Smith. [Massively multilingual word embeddings](#), February 2016.
- [365] Philip Massey, Patrick Xia, David Bamman, and Noah A. Smith. [Annotating character relationships in literary texts](#), December 2015.
- [366] 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.
- [367] 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.
- [368] 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.
- [369] Lingpeng Kong and Noah A. Smith. [An empirical comparison of parsing methods for stanford dependencies](#), April 2014.
- [370] Justin H. Gross, Brice D. L. 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.
- [371] David Bamman and Noah A. Smith. [New alignment methods for discriminative summarization](#), May 2013.
  - [372] David Bamman, Adam Anderson, and Noah A. Smith. [Inferring social rank in an Old Assyrian trade network](#). July 2013. Presented at Digital Humanities.
  - [373] 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.
  - [374] Noah A. Smith. [Adversarial evaluation for models of natural language](#), July 2012.
  - [375] 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.
  - [376] 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.
  - [377] 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.
  - [378] 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.
  - [379] 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.
  - [380] Noah A. Smith. [Text-driven forecasting](#), March 2010.
  - [381] 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.
  - [382] 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.
  - [383] 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. .
  - [384] 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 (selected)*

- [385] [SOPA](#), developed by Roy Schwartz and Sam Thomson. Soft patterns, 2018, see [110].
- [386] [SCHOLAR](#), developed by Dallas Card. Tool for modeling documents with metadata, 2018, see [107].
- [387] [IDEA\\_RELATIONS](#), developed by Chenhao Tan and Dallas Card. A framework to identify relations between ideas in temporal text corpora, 2016, see [121].
- [388] [TWITTER LANGID](#), developed by Aaron Jaech and Phoebe Mulcaire. Word-level language identification for tweets, 2016, see [272].
- [389] [RECURRENT NEURAL NETWORK GRAMMARS](#), developed by Chris Dyer, Miguel Ballesteros, Adhiguna Kuncoro. 2016, see [128].

- [390] [PORTAL FOR EVALUATING MULTILINGUAL WORD EMBEDDINGS](#), developed by Waleed Ammar. 2016, see [364].
- [391] [OPEN EXTRACTION OF FINE-GRAINED POLITICAL STATEMENTS](#), developed by David Bamman. 2015, see [133].
- [392] [STACK LSTM PARSER AND EXTENSIONS](#), developed by Chris Dyer, Miguel Ballesteros, Adhiguna Kuncoro, SwabhaSwayamdipta, and Waleed Ammar. 2015, see [20, 124, 127, 132, 136].
- [393] [CHARACTERRELATIONS](#), developed by Philip Massey, Patrick Xia, and David Bamman. Annotations of character relations in 109 literary texts, 2015, see [365].
- [394] [GFL-WEB](#), developed by Michael T. Mordowanec. Web interface for the Graph Fragment Language for syntactic annotation, 2014, see [246].
- [395] [JAMR](#), developed by Jeffrey Flanigan and Sam Thomson. Semantic parser and generator for the Abstract Meaning Representation, 2014–6, see [129, 151].
- [396] [COMPREHENSIVE MULTIWORD EXPRESSIONS CORPUS](#), developed by Nathan Schneider. English Web Treebank annotated with multiword expressions, 2014, see [154].
- [397] [AMALGR](#), developed by Nathan Schneider. Multiword expression identification tool, 2014, see [26].
- [398] [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 [278].
- [399] [CMU MOVIE SUMMARY CORPUS](#), developed by David Bamman and Brendan O’Connor. Corpus of movie plot summaries and associated metadata, 2013, see [158].
- [400] [GLOBAL VOICES MALAGASY-ENGLISH PARALLEL CORPUS](#), developed by Victor Chahuneau. Corpus of parallel news articles from the Global Voices citizen media project, 2012.
- [401] [WORD SALAD](#), developed by Victor Chahuneau. Corpus of restaurant menus, 2012, see [162].
- [402] [ARABIC WIKIPEDIA SUPERSENSE CORPUS](#), developed by Nathan Schneider and Behrang Mohit. Articles tagged with nominal supersenses, 2012, see [251].
- [403] [AD<sup>3</sup>](#), developed by André Martins. Approximate MAP decoder, 2012, see [21, 164, 171, 174, 291].
- [404] [RAMPION](#), developed by Kevin Gimpel. Algorithm for training statistical machine translation models based on minimizing structured ramp loss, 2012, see [166].
- [405] [CONGRESSIONAL BILLS CORPUS](#), developed by Tae Yano. Congressional bills and committee outcomes, 2012, see [167].
- [406] [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 [168].
- [407] [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 [161, 254].
- [408] [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 [300].
- [409] [AMAZON MECHANICAL TURK POLITICAL BIAS DATA](#), developed by Tae Yano. Sentences from political blogs with crowdsourced annotations of political bias, 2010, see [296].
- [410] [MOVIE\\$ CORPUS](#), developed by Mahesh Joshi, Dipanjan Das, and Kevin Gimpel. Collection of pre-release movie reviews, metadata, and opening weekend revenues, 2010, see [256].
- [411] [SEMAFOR](#), developed by Dipanjan Das, Nathan Schneider, and Desai Chen. Frame-semantic parser, 2010, see [186].
- [412] [TURBOPARSER](#), developed by André Martins. Multilingual dependency parser, 2009, see [192].

- [413] [QUIPU](#), developed by Kevin Gimpel. Statistical machine translation system, 2009, see [190].
- [414] [POLITICAL BLOG CORPUS](#), developed by Tae Yano. Text collection from five American political blogs, 2009, see [197].
- [415] [10-K CORPUS](#), developed with three others. Collection of 10-K reports and preceding and following stock return volatility measurements, 2009, see [195].
- [416] [DAGEEM](#), developed by Shay Cohen. Unsupervised dependency grammar induction, 2008, see [200].
- [417] [MSTPARSER](#), [STACKED](#), developed by André Martins and Dipanjan Das. Multilingual dependency parser, 2008, see [201].
- [418] [CURD](#), developed by Dan Tasse. Corpus of semantically annotated recipes, 2008, see [382].
- [419] [DYNA](#), developed with five others. Declarative programming language for weighted dynamic programming, 2004, see [211, 263].
- [420] [EGYPT](#), developed with nine others. Toolkit for statistical machine translation, including GIZA training module and CAIRO word alignment visualizer, 1999, see [264, 383].

## PATENTS

- “Entity-specific tuned searching.” Eric D. Brill, Noah A. Smith, and Robert J. Ragno. U.S. patent 7739270, assigned to Microsoft Corporation, June 15, 2010 (filed December 7, 2004).