

# Wei Xu

✉ weixu@cse.ohio-state.edu    ☎ +1 614 292 3083    🐦 @cocoweixu  
395 Dreese Laboratories, 2015 Neil Avenue Columbus, OH, U.S.A. 43210  
<http://web.cse.ohio-state.edu/~weixu/>

---

RESEARCH	Natural Language Processing, Machine Learning, Social Media	
EXPERIENCE	<b>Assistant Professor, Ohio State University</b> , Columbus, OH <i>Department of Computer Science and Engineering</i>	Aug 2016 – present
EDUCATION	<b>Ph.D. in Computer Science, New York University</b> , New York, NY Advisor: Ralph Grishman; Committee: Satoshi Sekine, Ernest Davis, Bill Dolan (Microsoft Research), Luke Zettlemoyer (University of Washington/Facebook AI)	2014
	<b>B.S. and M.S. in Computer Science, Tsinghua University</b> , Beijing, CHINA	2004/2007
EXPERIENCE	<b>Postdoctoral Researcher, University of Pennsylvania</b> , Philadelphia, PA <i>Computer Information and Science Department</i> (Advisor: Chris Callison-Burch)	Feb 2014 – Aug 2016
	<b>Visiting Graduate Student, University of Washington</b> , Seattle, WA <i>Computer Science and Engineering Department</i>	Jan 2012 – Dec 2013
	<b>Intern, Microsoft Research</b> , Seattle, WA <i>Internet Services Research Center</i>	Jun – Sep 2011
	<b>Intern, Educational Testing Service</b> , Princeton, NJ <i>Natural Language and Speech Group</i>	Mar – May, Sep – Dec 2010
	<b>Intern, Amazon</b> , Seattle, WA <i>Catalog Quality</i>	May – Sep 2010
AWARDS, GRANTS, GIFTS	<b>Best Paper Award, COLING</b> , 2018 <b>CrowdFlower AI for Everyone Award</b> , 2018 (\$25,000 PI) <b>Criteo Faculty Research Award</b> , 2018 (\$20,000 PI) <b>NSF CRII Award</b> , 2018 (\$175,000 PI) <i>Robust Intelligence: Learning a Timely Semantic Resource from Social Media Data</i> <b>NSF Grant</b> , 2018 – 2021 (\$375,732 PI) <i>Collaborative Research: Automatic Text-Simplification and Reading-Assistance to Support Self-Directed Learning by Deaf and Hard-of-Hearing Computing Workers</i> <b>DARPA Grant</b> , 2017 – 2021 (\$600,000 co-PI 50%) <i>Computational Simulation of Online Social Behavior</i> <b>NYU MacCracken PhD Fellowship</b> , 2007 – 2012 (\$230,000) <b>Ohio Supercomputer Center Computing Support</b> , 2018 <b>NVIDIA Academic Hardware Gift</b> , 2018 <b>Microsoft Research Gift for WNUT Workshop</b> , 2015	
SELECTED PUBLICATIONS	(Note: <u>Underline</u> is used to indicate student advisees at the Ohio State University.) <i>A Word-Complexity Lexicon and A Neural Readability Ranking Model for Lexical Simplification</i> <u>Mounica Maddela</u> , Wei Xu Proceedings of EMNLP 2018, long paper, oral presentation (acceptance rate 25.8%) <i>Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering</i> <u>Wuwei Lan</u> , Wei Xu Proceedings of COLING 2018, long paper ( <b>Best Paper Award</b> ; selection rate $8/888 = 0.90\%$ ) <i>Taking the First Essential Steps in Automating the Wet Laboratory: Annotating a Corpus of Protocols for Reproducibility</i> <u>Chaitanya Kulkarni</u> , Wei Xu, Alan Ritter, Raghu Machiraju NAACL 2018, short paper (acceptance rate 29%) <i>The Importance of Subword Embeddings in Sentence Pair Modeling</i> <u>Wuwei Lan</u> , Wei Xu NAACL 2018, short paper (acceptance rate 29%)	

*A Continuously Growing Dataset of Sentential Paraphrases*

Wuwei Lan, Siyu Qiu, Hua He, Wei Xu

EMNLP 2017, long paper (acceptance rate 25.8%)

*A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter*

Jeniya Tabassum, Alan Ritter, Wei Xu

EMNLP 2016, long paper (acceptance rate 26%)

*Optimizing Statistical Machine Translation for Simplification*

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch

TACL 2016, long paper, oral presentation at ACL 2016

*Discovering User Attribute Stylistic Differences via Paraphrasing*

Daniel Preotiuc, Wei Xu, Lyle Ungar

AAAI 2016, long paper, oral presentation (acceptance rate 26%)

*Results of the WNUT16 Named Entity Recognition Shared Task*

Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine de Marneffe, Wei Xu

COLING 2016 Workshop on Noisy User-generated Text (shared-task overview)

*Problems in Current Text Simplification Research: New Data Can Help*

Wei Xu, Chris Callison-Burch, Courtney Napoles

TACL 2015, long paper, oral presentation at EMNLP 2015

*Cost Optimization for Crowdsourcing Translation*

Mingkun Gao, Wei Xu, Chris Callison-Burch

NAACL 2015, long paper (acceptance rate 29%)

*SemEval-2015 Task 1: Paraphrase and Semantic Similarity in Twitter*

Wei Xu, Chris Callison-Burch, William B. Dolan

SemEval 2015 (shared-task overview)

*Extracting Lexically Divergent Paraphrases from Twitter*

Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, Yangfeng Ji

TACL 2014, long paper, oral presentation at NAACL 2015

*Infusion of Labeled Data into Distant Supervision for Relation Extraction*

Maria Pershina, Bonan Min, Wei Xu, Ralph Grishman

ACL 2014, short paper (acceptance rate 25.2%)

*Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction*

Wei Xu, Raphael Hoffmann, Le Zhao, Ralph Grishman

ACL 2013, short paper (acceptance rate 24%)

*Paraphrasing for Style*

Wei Xu, Alan Ritter, Bill Dolan, Ralph Grishman, Colin Cherry

COLING 2012, long paper (acceptance rate 25%)

*Exploiting Syntactic and Distributional Information for Spelling Correction with Web-Scale N-grams Models*

Wei Xu, Joel Tetreault, Martin Chodorow, Ralph Grishman, Le Zhao

EMNLP 2011, long paper (acceptance rate 23.7%)

*Passage Retrieval for Information Extraction using Distant Supervision*

Wei Xu, Ralph Grishman, Le Zhao

IJCNLP 2011, long paper

*Extractive Summarization using Inter- and Intra- Event Relevance*

Wenjie Li, Wei Xu, Mingli Wu, Chunfa Yuan, Qin Lu

ACL 2006, long paper, oral presentation (acceptance rate 23%)

## SERVICE

**Workshop Chair:** ACL (2017)

**Publicity Chair:** EMNLP (2019), NAACL (2018, 2016)

**Area Chair:** ACL (2019), NAACL (2019), COLING (2018), EMNLP (2018, 2016)

**NSF Review Panelist:** CISE (2017)

**Journal Reviewer:** TACL, JAIR

**Program Committee:** ACL (2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2018, 2015), WWW (2017, 2016, 2015), AAAI (2016, 2015, 2012), KDD (2015), COLING (2014)

**Organizer:** Workshop on Noisy User-generated Text (EMNLP 2018 & 2017, COLING 2016, ACL 2015); Shared-task on Paraphrase Identification and Semantic Similarity in Twitter (SemEval 2015)

<b>INVITED TALKS</b>	<i>Learning for Unlimited Human Language</i>	Dec 2018
	Peking University, Beijing, China	
	<i>Learning Large-scale Paraphrases for Natural Language Understanding and Generation</i>	Nov 2017 - 2018
	Midwest Machine Learning Symposium, Chicago; Facebook, Menlo Park; Twitter, San Francisco; IBM Thomas J. Watson Research Center, New York	
	<i>How AI Understand Language?</i>	Mar 2018
	Women in Analytics Conference, Columbus, OH	
	<i>Can Paraphrase be a Ultimate Solution for NLU and NLG?</i>	July 2017
	Google Research, New York	
	<i>Paraphrase <math>\approx</math> Monolingual Translation</i>	Aug 2016
	Amazon, Berlin, Germany	
<b>TEACHING</b>	<i>Multiple Instance Learning from Unlimited Text</i>	Oct 2015 – Dec 2016
	Microsoft Research Asia; University of Delaware; University of Edinburgh; University of North Carolina, Chapel Hill; Arizona State University; Vanderbilt University; Imperial College London; University of Waterloo; Indiana University, Bloomington; Washington University, St Louis; Simon Fraser University; University of Alberta; Yale University; University of Maryland	
	<i>Learning and Generating Paraphrases from Twitter and Beyond</i>	Feb 2015
	Carnegie Mellon University; Columbia University; Johns Hopkins University	
	<i>Paraphrases in Twitter</i>	Feb 2015
	Twitter, San Francisco	
	<i>Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!)</i>	Aug 2014 – Mar 2015
	City University of New York; IBM Research, Almaden; University of California, Berkeley; University of Texas, Austin; Yahoo!, New York; Carnegie Mellon University; Microsoft Research, Redmond	
	<i>CSE 5522 Artificial Intelligence II: Advanced Techniques</i>	Spring 2018, Autumn 2018
	Ohio State University, mixed undergraduate and graduate level (teaching evaluation: 4.50/5.00 Spring 2018, 4.85/5.00 Autumn 2018).	
<b>Outreach</b>	<i>CSE 5525 Speech and Language Processing</i>	Spring 2017
	Ohio State University, mixed undergraduate and graduate level (teaching evaluation 3.80/5.00).	
	<i>CSE 5539 Social Media and Text Analytics</i>	Summer 2016, Autumn 2016, Autumn 2017
	<a href="http://socialmedia-class.org/">http://socialmedia-class.org/</a>	
	Designed a new course integrated with research, covering from basic to state-of-the-art machine learning algorithms (teaching evaluation: 5.72/6.00 Summer 2015, 4.60/5.00 Autumn 2016, 4.40/5.00 Autumn 2017). The course material I created has been used by other universities (e.g., University of Utah CS 1060, Georgia Tech CS 6452, Iowa State University MIS 546).	
	<b>Panelist:</b> Women in Analytics Conference (2018), CogFest - Cognitive Science Festival (2018)	
	<b>Mentor:</b> Women and Underrepresented Minorities in NLP Workshop (2017, 2018)	
	<b>Judge:</b> HackOHIO (2018), OHIO High School Hackathon (2017)	
	<b>Instructor:</b> North American Summer School on Logic, Language, and Information (2016)	
	<b>Organizer:</b> Mid-Atlantic Student Colloquium on Speech, Language and Learning (2016)	
<b>Students</b>	<b>Current Ph.D. Advisees:</b> Wuwei Lan, Mounica Maddela, Chao Jiang, Jeniya Tabassum (2/4 are woman)	
	<b>Current Undergrad Advisees:</b> Lillian Chow, Sydney Lee, Rita Tong, Raleigh Potluri (3/4 are woman)	
	<b>Former Student Advisees:</b> Siyu Qiu, Mingkun Gao, Jim Chen, Ray Lei, Maria Pershina, Wenchao Du, Chaitanya Kulkarni, Piyush Ghai, Pravar Mahajan (2/9 are women)	
	<b>Chris Callison-Burch</b> (✉ <a href="mailto:ccb@cis.upenn.edu">ccb@cis.upenn.edu</a> )	
	Associate Professor, Computer and Information Science Department	
	University of Pennsylvania, Philadelphia, PA	
	Sloan Research Fellow (2014), Chair of the NAACL Executive Board (2012-2013)	
	<b>Raymond J. Mooney</b> (✉ <a href="mailto:mooney@cs.utexas.edu">mooney@cs.utexas.edu</a> )	
	Professor, Computer Science Department, University of Texas, Austin, TX	
	ACL fellow (2014), ACM fellow (2010), AAAI fellow (2005)	
<b>REFERENCES</b>	<b>Bill Dolan</b> (✉ <a href="mailto:billdol@microsoft.com">billdol@microsoft.com</a> )	
	Research Manager/Principal Researcher, Microsoft Research, Redmond, WA	
	Manager of the Natural Language Processing Group	

**Joel Tetreault** (✉ [joel.tetreault@grammarly.com](mailto:joel.tetreault@grammarly.com))

Director of Research, Grammarly, New York, NY

Treasurer of NAACL (2013-2016)

**Ralph Grishman** (✉ [grishman@cs.nyu.edu](mailto:grishman@cs.nyu.edu))

Professor, Computer Science Department, New York University, New York, NY

ACL Fellow (2017), President of ACL (1991), Department Chairman (1986-1988)