

Wei Xu

✉ weixu@cse.ohio-state.edu ☎ +1 718 612 9347
395 Dreese Laboratories, 2015 Neil Avenue Columbus, OH, U.S.A. 43210
<http://web.cse.ohio-state.edu/~weixu/>

POSITION	Assistant Professor, Ohio State University , Columbus, OH <i>Department of Computer Science and Engineering</i> Research Areas: Natural Language Processing, Machine Learning, Social Media	Aug 2016 – present
EDUCATION	Ph.D. in Computer Science, New York University , New York, NY Thesis: <i>Data-driven Approaches for Paraphrasing Across Language Variations</i> (Advisor: Ralph Grishman) B.S. and M.S. in Computer Science, Tsinghua University , Beijing, CHINA Thesis: <i>Event-relevance Based Summarization</i>	2014 2004/2007
EXPERIENCE	Postdoctoral Researcher, University of Pennsylvania , Philadelphia, PA <i>Computer Information and Science Department</i> (Advisor: Chris Callison-Burch) Visiting Researcher, University of Washington , Seattle, WA <i>Computer Science and Engineering Department</i> Research Intern, Microsoft Research , Seattle, WA <i>Internet Services Research Center</i> Research Consultant, Educational Testing Service , Princeton, NJ <i>Natural Language and Speech Group</i>	Feb 2014 – Aug 2016 Jan 2012 – Dec 2013 Jun – Sep 2011 Mar – Dec 2010
GRANTS & AWARDS	CrowdFlower AI for Everyone Award (sole PI: ~\$150,000) DARPA Grant Computational Simulation of Online Social Behavior (co-PI 50% of \$600,000) NYU MacCracken Fellowship (total \$230,000 award)	2018 2017 – 2021 2007 – 2012
SERVICE	Workshop Chair: ACL (2017); Publicity Chair: NAACL (2016) Area Chair: COLING (2018), EMNLP (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)	
SELECTED PUBLICATIONS	<i>A Continuously Growing Dataset of Sentential Paraphrases</i> Wuwei Lan, Siyu Qiu, Hua He, Wei Xu Proceedings of EMNLP 2017, long paper (acceptance rate 25.8%) <i>A Minimally Supervised Method for Recognizing and Normalizing Time Expressions in Twitter</i> Jeniya Tabassum, Alan Ritter, Wei Xu Proceedings of EMNLP 2016, long paper (acceptance rate 26%), oral presentation <i>Optimizing Statistical Machine Translation for Simplification</i> Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, Chris Callison-Burch TACL 2016; oral presentation at ACL 2016 <i>Discovering User Attribute Stylistic Differences via Paraphrasing</i> Daniel Preotiuc, Wei Xu, Lyle Ungar Proceedings of AAAI 2016, long paper (acceptance rate 26%) <i>Problems in Current Text Simplification Research: New Data Can Help</i> Wei Xu, Chris Callison-Burch, Courtney Napoles TACL 2015; oral presentation at EMNLP 2015 <i>Cost Optimization for Crowdsourcing Translation</i> Mingkun Gao, Wei Xu, Chris Callison-Burch Proceedings of NAACL 2015, long paper (acceptance rate 29%) <i>SemEval-2015 Task 1: Paraphrase and Semantic Similarity in Twitter</i> Wei Xu, Chris Callison-Burch, William B. Dolan Proceedings of SemEval 2015 <i>Extracting Lexically Divergent Paraphrases from Twitter</i> Wei Xu, Alan Ritter, Chris Callison-Burch, William B. Dolan, Yangfeng Ji TACL 2014; oral presentation at NAACL 2015	

Infusion of Labeled Data into Distant Supervision for Relation Extraction

Maria Pershina, Bonan Min, Wei Xu, Ralph Grishman

Proceedings of 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

Proceedings of ACL 2013, short paper (acceptance rate 24%)

Gathering and Generating Paraphrases from Twitter with Application to Normalization

Wei Xu, Alan Ritter, Ralph Grishman

Proceedings of ACL Workshop on Building and Using Comparable Corpora 2013

Paraphrasing for Style

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

Proceedings of 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

Proceedings of EMNLP 2011, long paper (acceptance rate 23.7%)

Passage Retrieval for Information Extraction using Distant Supervision

Wei Xu, Ralph Grishman, Le Zhao

Proceedings of IJCNLP 2011, long paper

Extractive Summarization using Inter- and Intra- Event Relevance

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

Proceedings of COLING-ACL 2006, long paper, oral presentation (acceptance rate 23%)

**SELECTED
OPEN SOURCE
CODE & DATA**

A Continuously Growing Dataset of Sentential Paraphrases

Sep 2017

<https://github.com/lanwuwei/paraphrase-dataset/>

Text Simplification System and Newsela Corpus

May 2015

<https://github.com/cocoxu/simplification/>; <https://newsela.com/data/>

Multi-instance Learning Paraphrase Model and SemEval Twitter Paraphrase Corpus

Dec 2014

<https://www.cis.upenn.edu/~xwe/multi/>; <http://alt.qcri.org/semeval2015/task1/>

Parallel Shakespeare Corpus and Model

Jul 2012

<https://github.com/cocoxu/Shakespeare/>

**STUDENT
ADVISING**

Wuwei Lan (3rd year PhD student at OSU)

Mounica Maddela (1st year PhD student at OSU)

Shuaichen Chang (1st year PhD student at OSU)

Pravar Mahajan (masters student at OSU)

Piyush Ghai (masters student at OSU)

**INVITED
TALKS**

Learning Large-scale Paraphrases for Natural Language Understanding and Generation

Nov 2017

IBM Thomas J. Watson Research Center, New York

Can Paraphrase be a Ultimate Solution for NLU and NLG?

July 2017

Google Research, New York

Paraphrase \approx Monolingual Translation

Aug 2016

Amazon.com, Berlin, Germany

Multiple Instance Learning from Unlimited Text

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

Learning and Generating Paraphrases from Twitter and Beyond

Feb 2015

Carnegie Mellon University; Columbia University; Johns Hopkins University

Paraphrases in Twitter

Feb 2015

Twitter.com, San Francisco

Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!)

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

REFERENCES

Chris Callison-Burch

Associate Professor, Computer and Information Science Department
University of Pennsylvania, Philadelphia, PA
Sloan Research Fellow (2014), Chair of the NAACL Executive Board (2012-2013)
☎ +1 267 909 2668
✉ ccb@cis.upenn.edu
<http://www.cis.upenn.edu/~ccb/>

Raymond J. Mooney

Professor, Computer Science Department, University of Texas, Austin, TX
ACL fellow (2014), ACM fellow (2010), AAAI fellow (2005)
☎ +1 512 471 9558
✉ mooney@cs.utexas.edu
<https://www.cs.utexas.edu/~mooney/>

Ralph Grishman

Professor, Computer Science Department, New York University, New York, NY
ACL Fellow (2017), President of ACL (1991), Department Chairman (1986-1988)
☎ +1 212 998 3497
✉ grishman@cs.nyu.edu
<http://cs.nyu.edu/grishman/>

Bill Dolan

Research Manager/Principal Researcher, Microsoft Research, Redmond, WA
Manager of the Natural Language Processing Group
☎ +1 425 706 3709
✉ billdol@microsoft.com
<http://research.microsoft.com/en-us/people/billdol/>

Joel Tetreault

Director of Research, Grammarly, New York, NY
Treasurer of NAACL (2013-2016)
☎ +1 310 612 0315
✉ joel.tetreault@grammarly.com
<http://www.cs.rochester.edu/~tetreaul/academic.html>