Wei Xu

wei.xu@cc.gatech.edu https://cocoxu.github.io/ @cocoweixu
CODA Tech Square, Atlanta, GA, 30308 U.S.A.

Research Natural Language Processing, Machine Learning, Social Media

Citizenship United States

Academic Appointments

Assistant Professor, Georgia Institute of Technology, Atlanta, GA — Aug 2020 – Present
College of Computing, School of Interactive Computing

Adjunct Assistant Professor, The Ohio State University, Columbus, OH — Aug 2020 – July 2021
Department of Computer Science and Engineering

Assistant Professor, The Ohio State University, Columbus, OH — Aug 2016 – July 2020
Department of Computer Science and Engineering

Visiting Faculty, Carnegie Mellon University, Pittsburgh, PA — Summer 2019
Language Technologies Institute (Host: Graham Neubig)

Postdoctoral Researcher, University of Pennsylvania, Philadelphia, PA — Feb 2014 – Aug 2016
Computer Information and Science Department (Advisor: Chris Callison-Burch)

Visiting PhD Student, University of Washington, Seattle, WA — Jan 2012 – Dec 2013
Computer Science and Engineering Department

Education

Ph.D. in Computer Science, New York University, New York, NY — 2014
Advisor: Ralph Grishman; Committee: Satoshi Sekine, Ernest Davis, Bill Dolan (Microsoft Research), Luke Zettlemoyer (University of Washington/Facebook AI Research)

B.S./M.S. in Computer Science, Tsinghua University, Beijing, CHINA — 2004/2007

Grants

IARPA Grant *Human Interpretable Attribution of Text Using Underlying Structure* — 2022 – 2026
co-PI (50%), total $1,104,000

NSF CAREER *An Integrated Framework for Controllable Text Generation* — 2022 – 2027
PI (100%), total $537,527

NSF AI Institute *Collaborative Assistance and Responsive Interaction for Networked Groups* — 2021 – 2026
Senior Personnel, total $19,995,808

IARPA Research Grant *Better Extraction from Text Towards Enhanced Retrieval* — 2019 – 2023
co-PI (50%), total $850,000

NSF Grant *Collaborative Research: Automatic Text-Simplification and Reading-Assistance to Support Self-Directed Learning by Deaf and Hard-of-Hearing Computing Workers* — 2018 – 2022
PI (100%), total $375,732

NSF CRII *RI: Learning a Timely Semantic Resource from Social Media Data* — 2018 – 2021
PI (100%), total $183,000

DARPA Research Grant *Computational Simulation of Online Social Behavior* — 2017 – 2021
co-PI (50%), total $600,000

Awards

NSF CAREER Award, 2022

NSF CRII Award, 2018

Best Paper Award, **COLING**, 2018

Criteo Faculty Research Award, 2018

CrowdFlower AI for Everyone Award, 2018

NYU MacCracken PhD Fellowship, 2007 – 2012

Professional Services

Best Paper Award Committee: EMNLP (2018); **Senior Area Chair:** EMNLP (2022), NAACL (2022, 2021), ACL (2020); **Area Chair:** EMNLP (2021, 2020, 2018, 2016), AAAI (2020), ACL (2019), NAACL (2019), COLING (2018); **Action Editor:** ACL Rolling Review (2021-2022); **Program Committee:** ACL (2021, 2018, 2017, 2015, 2014, 2013), EMNLP (2017, 2015, 2014), NAACL (2015), WWW (2017, 2016, 2015), AAAI (2016, 2015, 2012), KDD (2015), COLING (2014); **Publicity Chair:** EMNLP (2019), NAACL (2018, 2016); **Workshop Chair:** ACL (2017); **Journal Reviewer:** Transactions of the Association for Computational (TACL), Journal of Artificial Intelligence Research (JAIR); **Organizer:** Workshop on Text Simplification, Accessibility, and Readability at EMNLP (2022); Workshop on Natural Language Generation, Evaluation, and Metrics at ACL (2021); Workshop on Noisy User-generated Text (`http://noisy-text.github.io/`) at EMNLP (2021, 2020, 2019, 2018, 2017), COLING (2022, 2016), ACL (2015); Mid-Atlantic Student Colloquium on Speech, Language and Learning (2016).

Publications (Underline is used to indicate student advisees.)

Improving Large-scale Paraphrase Acquisition and Generation
Yao Dou, Chao Jiang, Wei Xu
EMNLP 2022, long paper

Understanding the Human Revision Process in Scientific Writing
Chao Jiang, Wei Xu, Sam Stevens
EMNLP 2022, long paper

Stanceosaurus: Classifying Stance Towards Multicultural Misinformation
Jonathan Zheng, Ashutosh Baheti, Tarek Naous, Wei Xu, Alan Ritter
EMNLP 2022, long paper

A Dataset of Word-Complexity Judgements from Deaf and Hard-of-Hearing Adults for Text Simplification
Oliver Alonzo, Sooyeon Lee, Mounica Maddela, Wei Xu and Matt Huenerfauth
EMNLP TSAR Workshop 2022, long paper

BiSECT: Learning to Split and Rephrase Sentences with Bitexts
Joongwon Kim*, Mounica Maddela*,Reno Kriz, Wei Xu, Chris Callison-Burch (*equal contribution)
EMNLP 2021, long paper (acceptance rate 23.3%)

Pre-train or Annotate? Domain Adaptation with a Constrained Budget
Fan Bai, Alan Ritter, Wei Xu
EMNLP 2021 (acceptance rate 23.3%)

WIKIBIAS: Detecting Multi-Span Subjective Biases in Language
Yang Zhong, Jingfeng Yang, Wei Xu, Diyi Yang
EMNLP 2021 (Findings)

Neural semi-Markov CRF for Monolingual Word Alignment
Wuwei Lan*, Chao Jiang*, Wei Xu (*equal contribution)
ACL 2021, long paper (acceptance rate 21.2%)

Controllable Text Simplification with Explicit Paraphrasing
Mounica Maddela, Fernando Alva-Manchego, Wei Xu
NAACL 2021, long paper (acceptance rate 28%)

The GEM Benchmark: Natural Language Generation, its Evaluation and Metrics
Sebastian Gehrmann, Tosin Adewumi, Karmanya Aggarwal, Pawan Sasanka Ammanamanchi, Anuoluwapo Aremu, Antoine Bosselut, Khyathi Raghavi Chandu, Miruna-Adriana Clinciu, Dipanjan Das, Kaustubh Dhole, Wanyu Du, Esin Durmus, OndÅŹej DuÅąek, Chris Chinenye Emezue, Varun Gangal, Cristina Garbacea, Tatsunori Hashimoto, Yufang Hou, Yacine Jernite, Harsh Jhamtani, Yangfeng Ji, Shailza Jolly, Mihir Kale, Dhruv Kumar, Faisal Ladhak, Aman Madaan, Mounica Maddela, Khyati Mahajan, Saad Mahamood, Bodhisattwa Prasad Majumder, Pedro Henrique Martins, Angelina McMillan-Major, Simon Mille, Emiel van Miltenburg, Moin Nadeem, Shashi Narayan, Vitaly Nikolaev, Andre Niyongabo Rubungo, Salomey Osei, Ankur Parikh, Laura Perez-Beltrachini, Niranjan Ramesh Rao, Vikas Raunak, Juan Diego Rodriguez, Sashank Santhanam, JoÃčo Sedoc, Thibault Sellam, Samira Shaikh, Anastasia Shimorina, Marco Antonio Sobrevilla Cabezudo, Hendrik Strobelt, Nishant Subramani, Wei Xu, Diyi Yang, Akhila Yerukola, Jiawei Zhou (alphabetically ordered)
ACL 2021 Workshop on Natural Language Generation, Evaluation, and Metrics
(project website: `https://gem-benchmark.com/`)

An Empirical Study of Pre-trained Transformers for Arabic Information Extraction
Wuwei Lan, Yang Chen, Wei Xu, Alan Ritter
EMNLP 2020, short paper (acceptance rate 16.7%)

WNUT-2020 Task 1 Overview: Extracting Entities and Relations from Wet Lab Protocols
Jeniya Tabassum, Sydney Lee, Wei Xu, Alan Ritter
EMNLP 2020 Workshop on Noisy User-generated Text (shared-task overview)

Neural CRF Model for Sentence Alignment in Text Simplification
Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, Wei Xu
ACL 2020, long paper (acceptance rate 25.2%)

An Empirical Study of Named Entity Recognition in StackOverflow
Jeniya Tabassum, Mounica Maddela, Wei Xu, Alan Ritter
ACL 2020, long paper (acceptance rate 25.2%)

Generalizing Natural Language Analysis through Span-relation Representations
Zhengbao Jiang, Wei Xu, Jun Araki, Graham Neubig
ACL 2020, long paper (acceptance rate 25.2%)

Learning Relation Entailment with Structured and Textual Information
Zhengbao Jiang, Jun Araki, Donghan Yu, Ruohong Zhang, Wei Xu, Yiming Yang, Graham Neubig
AKBC 2020, long paper

Discourse Level Factors for Sentence Deletion in Text Simplification
Yang Zhong, Chao Jiang, Wei Xu, Junyi Jessy Li
AAAI 2020, long paper (acceptance rate 20.6%; oral presentation)

Multi-task Pairwise Neural Ranking for Hashtag Segmentation
Mounica Maddela, Wei Xu, Daniel Preotiuc-Pietro
ACL 2019, long paper (acceptance rate 25.7%)

A Word-Complexity Lexicon and A Neural Readability Ranking Model for Lexical Simplification
Mounica Maddela, Wei Xu
EMNLP 2018, long paper (acceptance rate 25.8%; oral presentation)

Neural Network Models for Paraphrase Identification, Semantic Textual Similarity, Natural Language Inference, and Question Answering
Wuwei Lan, Wei Xu
COLING 2018, long paper (**Best Paper Award**; selection rate 8/888 = 0.90%)

An Annotated Corpus for Machine Reading of Instructions in Wet Lab Protocols
Chaitanya Kulkarni, Wei Xu, Alan Ritter, Raghu Machiraju
NAACL 2018, short paper (acceptance rate 29%)

Character-based Neural Networks for Sentence Pair Modeling
Wuwei Lan, 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%)

From Shakespeare to Twitter: What are Language Styles all about?
Wei Xu
EMNLP 2017 Workshop on Stylistic Variation

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%; oral presentation)

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-Pietro, Wei Xu, Lyle Ungar
AAAI 2016, long paper (acceptance rate 26%; oral presentation)

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, long paper (shared-task overview)

Shared Tasks of the 2015 Workshop on Noisy User-generated Text: Twitter Lexical Normalization and Named Entity Recognition
Timothy Baldwin, Marie Catherine de Marneffe, Bo Han, Young-Bum Kim, Alan Ritter, Wei Xu
ACL 2015 Workshop on Noisy User-generated Text (shared-task overview; author ordered alphabetically)

Data-driven Approaches for Paraphrasing Across Language Variations
Wei Xu
Ph.D. Thesis 2014

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%; oral presentation)

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%)

Gathering and Generating Paraphrases from Twitter with Application to Normalization
Wei Xu, Alan Ritter, Ralph Grishman
ACL 2013 Workshop on Building and Using Comparable Corpora

A Preliminary Study of Tweet Summarization using Information Extraction
Wei Xu, Ralph Grishman, Adam Meyers, Alan Ritter
NAACL 2013 Workshop on Language Analysis in Social Media

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%)

New York University 2011 System for KBP (Knowledge Base Population) Slot Filing
Ang Sun, Ralph Grishman, Wei Xu, Bonan Min
TAC 2011 (best performance system in NIST KBP-2011 evaluation)

Passage Retrieval for Information Extraction using Distant Supervision
Wei Xu, Ralph Grishman, Le Zhao
IJCNLP 2011, long paper (acceptance rate 36%)

Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task
Kristen Parton, Kathleen McKeown, Bob Coyne, Mona Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Weiyun Ma, Adam Meyers, Sara Stolbach, Ang Sun, Gokhan Tur, Wei Xu, Sibel Yaman
ACL 2009, long paper (acceptance rate 21%; oral presentation)

A Parse-and-Trim Approach with Information Significance for Chinese Sentence Compression
Wei Xu, Ralph Grishman
ACL Workshop on Language Generation and Summarisation 2009

Transducing Logical Relations from Automatic and Manual Annotation
Adam Meyers, Michiko Kosaka, Heng Ji, Nianwen Xue, Mary Harper, Ang Sun, Wei Xu, Shasha Liao
ACL Workshop on Linguistic Annotation 2009

Automatic Recognition of Logical Relations for English, Chinese and Japanese in the GLARF Framework
Adam Meyers, Michiko Kosaka, Nianwen Xue, Heng Ji, Ang Sun, Shasha Liao, Wei Xu
SemEval 2009, long paper

Extractive Summarization using Inter- and Intra- Event Relevance
Wenjie Li, Wei Xu, Mingli Wu, Chunfa Yuan, Qin Lu
ACL 2006, long paper (acceptance rate 23%; oral presentation)

Using Non-Local Features to Improve Named Entity Recognition Recall
Xinnian Mao, Wei Xu, Yuan Dong, Haila Wang
PACLIC 2007, long paper

Deriving Event Relevance from the Ontology Constructed with Formal Concept Analysis
Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan
CICLing 2006, long paper (acceptance rate 30.4%; oral presentation)

Building Document Graph for Text Summarization: An Event-based Approach
Wei Xu, Wenjie Li, Mingli Wu, Wei Li, Chunfa Yuan
ICCPOL 2006

The THU/PolyU System at MSE 2006: An Event-relevance based Approach
Wei Xu, Chunfa Yuan, Mingling Wu, Wenjie Li
MSE 2006

Teaching *CS 4650 Natural Language Processing*

Georgia Tech, undergraduate level (Spring 2021, Spring 2022)

CS 7650 Natural Language Processing
Georgia Tech, graduate level (Fall 2022, Fall 2021)

CSE 5539 Social Media and Text Analytics (`http://socialmedia-class.org/`)
A new course integrated with research, covering from basic to state-of-the-art machine learning algorithms
(teach eval: 4.13/5.00 Fall 2019, 4.40/5.00 Fall 2017, 4.60/5.00 Fall 2016; 5.72/6.00 at NASSLLI 2015)

CSE 5522 Artificial Intelligence II: Advanced Techniques
mixed undergraduate and graduate level
(teaching eval: 4.85/5.00 Fall 2018, 4.50/5.00 Spring 2018)

CSE 5525 Speech and Language Processing
mixed undergraduate and graduate level (teaching eval: 4.42/5.00 Spring 2020, 3.80/5.00 Spring 2017)

Students

Mounica Maddela (PhD student at GaTech; intern at Meta AI and Bloomberg AI)	2017 – present
Chao Jiang (PhD student at GaTech)	2018 – present
Yang Chen (PhD student at GaTech; intern at Google Research - co-advisor: Alan Ritter)	2019 – present
Yao Dou (PhD student at GaTech; intern at AI2)	2021 – present
Tarek Naous (PhD student at GaTech)	2022 – present
Duong Minh Le (PhD student at GaTech)	2022 – present
Mohit Chandra (PhD student at GaTech)	2022 – present
Junmo Kang (PhD student at GaTech - co-advisor: Alan Ritter)	2022 – present
Chase Perry (MS student at GaTech)	Summer 2022 – present
Jonathan Zheng (undergraduate at GaTech)	Fall 2020 – present
David Heineman (undergraduate at GaTech)	Winter 2020 – present
Michael Ryan (undergraduate at GaTech)	Winter 2020 – present
Vishnu Suresh (undergraduate at GaTech)	Fall 2021 – present
Elizabeth Liu (undergraduate at GaTech)	Spring 2022 – present
Marcus Ma (undergraduate at GaTech)	Spring 2022 – present
Vishnesh Jayanthi (undergraduate at GaTech)	Summer 2022 – present
Rachel Choi (undergraduate at GaTech)	Summer 2022 – present
Ian Ligon (undergraduate at GaTech)	Summer 2022 – present
Kelly Smith (undergraduate at GaTech)	Fall 2022 – present
Anton Lavrouk (undergraduate at GaTech)	Fall 2022 – present
Vinayak Athavale (undergraduate at GaTech)	Fall 2022 – present
Srushti Nandu (research intern at GaTech)	Summer 2021 – present
Alexandra Soong (undergraduate at GaTech)	Spring 2022
Andrew Duffy (research intern at GaTech)	Summer 2021 – Spring 2022
Wuwei Lan (completed PhD student at OSU, now Applied Scientist at Amazon)	2016 – 2021
Jeniya Tabassum (completed PhD student at OSU, now at Amazon - co-advisor: Alan Ritter)	2016 – 2021
Yang Zhong (completed MS student at OSU, now PhD student at UPitt)	2019 – 2021
Joong Kim (completed undergraduate at UPenn, now PhD student at UW)	Fall 2020 – Spring 2021
Piyush Ghai (completed MS student at OSU, now at Amazon)	Fall 2017
Ema Goh (undergraduate at GaTech)	2020 – 2021
Sydney Lee (completed undergraduate at OSU, now at Capital One)	2018 – 2020
Sarah Flanagan (completed undergraduate at OSU)	2018 – 2020
Sam Stevens (completed undergraduate at OSU)	2019 – 2020
Daniel Szoke (undergraduate at OSU)	2019 – 2020
Brian Seeds (undergraduate at OSU)	Summer 2020
Kenneth Koepcke (undergraduate at UIUC)	2020 – 2021
Muji Lai (high school intern at GaTech)	Spring 2022
Panya Bhinder (high school intern at OSU)	Summer 2020
Solomon Wood (high school intern at OSU)	Spring 2020
Raleigh Potluri (undergraduate student at OSU)	2018 – 2019
Lillian Chow (undergraduate student at OSU)	2018 – 2019
Rita Tong (completed undergraduate student at OSU, now MS student at UWisconsin)	2018 – 2019
Wenchao Du (completed undergraduate student at UWaterloo, now MS student at CMU LTI)	2016
Mingkun Gao (completed MS student at UPenn, now PhD student at UIUC)	2015
Siyu Qiu (completed MS student at UPenn, now at Hulu)	2015
Jim Chen (completed undergraduate student at UPenn, now PhD student at UW)	2014 – 2015
Ray Lei (completed undergraduate student at UPenn, now at Microsoft)	2014

Thesis Committee

Sarah Wiegreffe (PhD student at GaTech – advisor: Mark Riedl)	2022

	Yuval Pinter (completed PhD student at GaTech – advisor: Jacob Eisenstein)	2021
	Shi Zong (completed PhD student at OSU – advisor: Alan Ritter)	2021
	Sanqiang Zhao (completed PhD student at UPitt – advisor: Daqing He)	2021
	Kai Cao (completed PhD student at NYU – advisor: Ralph Grishman)	2017
	Maria Pershina (completed PhD student at NYU – advisor: Ralph Grishman)	2014

Invited Talks

Tracking the Spread of Multicultural (mis)Information

Columbia University, New York, NY	Oct 2022

Importance of Data and Controllability in Neural Language Generation

Cornell Tech (LMSS Seminar), New York, NY	Sep 2022
Dataminr, New York, NY	Nov 2021
Nanjing University, Nanjing, China	Oct 2021
SimpleText workshop at CLEF 2021	Sep 2021
Stanford University (NLP Seminar), Stanford, CA	Aug 2021
University of California, Los Angeles (Big Data and ML Seminar)	Jun 2021

Importance of Data and Linguistics in Neural Language Generation

New York University, New York, NY (NLP and Text-as-Data Speaker Series)	May 2021
Carnegie Mellon University, Pittsburgh, PA (LTI Colloquium)	Nov 2020
Google Research	Oct 2020

Natural Language Understanding for Noisy Text

University of Sheffield, Sheffield, United Kingdom (NLP Seminar)	Oct 2020
USC Information Sciences Institute, Los Angeles, CA (NLP Seminar)	Oct 2020

Automatic Text Simplification

University of Pittsburgh, Pittsburgh, PA (NLP Seminar)	Oct 2020

Understanding and Generating Human Language

Emory University, Atlanta, GA (CS Department Seminar)	Sep 2020
University of Maryland, College Park, MD (CS Colloquium)	Feb 2020
University of Massachusetts, Amherst, MA	Jan 2020
Georgia Institute of Technology, Atlanta, GA	Dec 2019

Learning for Unlimited Human Language

Peking University, Beijing, China	Dec 2018

Learning Large-scale Paraphrases for Natural Language Understanding and Generation

Midwest Machine Learning Symposium, Chicago, IL	Jun 2018
Facebook, Menlo Park, CA	May 2018
Stanford Research Institute, Menlo Park, CA	May 2018
Twitter, San Francisco, CA	May 2018
IBM Thomas J. Watson Research Center, New York, NY	Nov 2017

How does AI Understand Language?

Women in Analytics Conference, Columbus, OH (Main Stage Panel)	Mar 2018

Can Paraphrase be a Ultimate Solution for NLU and NLG?

Google Research, New York, NY	Jul 2017

Paraphrase ≈ Monolingual Translation

Amazon, Berlin, Germany	Aug 2016

Multiple Instance Learning from Unlimited Text

Microsoft Research Asia, Beijing, China	Dec 2016
University of Delaware, Newark, DE	Sep 2016
University of Edinburgh, Edinburgh, United Kingdom	May 2016
Ohio State University, Columbus, OH	Apr 2016
University of North Carolina, Chapel Hill, NC	Apr 2016
Arizona State University, Tempe, AZ	Mar 2016
Vanderbilt University, Nashville, TN	Mar 2016
Imperial College London, London, United Kingdom	Mar 2016
University of Waterloo, Waterloo, ON, Canada (CS Seminar)	Mar 2016
Indiana University, Bloomington, IN (Computer Science Colloquium Series)	Feb 2016
Washington University, St Louis, MI (Computer Science & Engineering Colloquia Series)	Feb 2016
Simon Fraser University, Vancouver, BC, Canada	Feb 2016
University of Alberta, Edmonton, AB, Canada	Feb 2016
Yale University, New Haven, CT (CS Talk)	Feb 2016
University of Maryland, College Park, MD (CLIP Colloquium)	Oct 2015

	Ohio State University, Columbus, OH (Clippers Seminar)	Oct 2015
	Large-scale Paraphrase Acquisition from Twitter	
	DARPA's DEFT Project Meeting, Boulder, CO	May 2015
	Learning and Generating Paraphrases from Twitter and Beyond	
	Carnegie Mellon University, Pittsburgh, PA	Apr 2015
	Columbia University, New York, NY (NLP Talk)	Apr 2015
	Johns Hopkins University, Baltimore, MD (CLSP Colloquium)	Feb 2015
	Paraphrases in Twitter	
	Twitter, San Francisco, CA	Feb 2015
	Modeling Lexically Divergent Paraphrases in Twitter (and Shakespeare!)	
	The City University of New York, New York, NY (NLP Seminar)	Mar 2015
	IBM Research - Almaden, San Jose, CA	Feb 2015
	University of California, Berkeley, CA	Feb 2015
	The University of Texas, Austin, TX (Forum for Artificial Intelligence)	Feb 2015
	Yahoo!, New York, NY	Dec 2014
	Carnegie Mellon University, Pittsburgh, PA (CL+NLP Lunch Seminar)	Nov 2014
	Microsoft Research, Seattle, WA (Visiting Speaker Series)	Aug 2014
	Incremental Information Extraction	
	Stanford Research Institute, Palo Alto, CA	Apr 2012
	IARPA's KDD Project Meeting, San Diego, CA	May 2011
	Event-based Summarization	
	Thomson Reuters, Eagan, Minnesota, MN	Nov 2009
Outreach Activities	Mentor, Group Mentoring Sessions for undergraduate/master students at ACL 2020	July 2020
	Speaker/Judge, Ohio High School Hackathon	Mar 2019
	Speaker, Franklin Friday art and science festival in Columbus Ohio	Mar 2019
	Panelist, CogFest - Cognitive Science Festival	Apr 2018
	Mentor, Women and Underrepresented Minorities in NLP Workshop	Jun 2018
	Mentor, OSU's AI Hackathon	Apr 2018
	Speaker/Panelist, Women in Analytics Conference	Mar 2018
	Speaker, OSU's AI Club	Feb 2018
	Judge, HackOhio	Oct 2017
	Mentor, Women and Underrepresented Minorities in NLP Workshop	Jul 2017
	Judge, Ohio High School Hackathon	Mar 2017
	Presenter, Philadelphia Science Festival	Apr 2015
Open Source Code / Data	*#HashtagMaster: A Semantic Analysis Tool for Hashtags*	Jun 2019
	`https://mounicam.github.io/hashtag_master`	
	Pairwise Neural Ranking Model and SimplePPDB++	Oct 2018
	`https://github.com/lanwuwei/SPM_toolkit`	
	SPM Toolkit for Sentence Pair Modeling	Aug 2018
	`https://github.com/lanwuwei/SPM_toolkit`	
	LanguageNet: Large-scale Paraphrase Corpus	Sep 2017
	`https://github.com/lanwuwei/paraphrase-dataset`	
	Syntax MT-based Text Simplification System and SARI Evaluation Metric	May 2015
	`https://github.com/cocoxu/simplification/` (contribution to the Joshua Machine Translation Toolkit)	
	NEWSELA Text Simplification Corpus	Sep 2015
	`https://newsela.com/data/` (widely adopted as the benchmark for text simplification research)	
	Multiple-instance Learning Paraphrase Model	Dec 2014
	`https://github.com/cocoxu/multip`	
	Twitter Paraphrase Corpus (shared-task at SemEval-2015)	Oct 2014
	`http://alt.qcri.org/semeval2015/task1/`	
	Event-based Twitter Summarization System	Nov 2013
	`https://github.com/cocoxu/twittersummarization/`	
	Twitter Normalization Phrase Table	Oct 2014
	`https://github.com/cocoxu/twitterparaphrase/`	
	Parallel Shakespeare Corpus and Model	Jul 2012
	`https://github.com/cocoxu/Shakespeare/`	