

Justine Zhang

<http://tisjune.github.io/>

jz727@cornell.edu

Research Areas

Natural Language Processing, Computational Social Science

Education

PhD in Information Science

Cornell University, Ithaca, NY, USA

Minor in Computer Science

Advised by Cristian Danescu-Niculescu-Mizil

Committee: Jon Kleinberg, Lillian Lee

GPA: 4.105/4

2015–2021

(expected)

BSc in Computer Science

Stanford University, Stanford, CA, USA

GPA: 3.93/4

2011–2015

Awards

Top reviewer recognition

ACL 2020, EMNLP 2018 and 2019, ICWSM 2020, NAACL 2019

EECS Rising Star

2020

Microsoft PhD Fellowship

2018–2020

Cornell Information Science Fellowship

2015

Publications

** indicates equal contribution*

- CSCW 2020a Quantifying the Causal Effects of Conversational Tendencies
Justine Zhang, Sendhil Mullainathan, and Cristian Danescu-Niculescu-Mizil.
Proceedings of CSCW, 2020. [\[link\]](#)
- ACL 2020 Balancing Objectives in Counseling Conversations: Advancing Forwards or Looking Backwards
Justine Zhang and Cristian Danescu-Niculescu-Mizil.
Proceedings of ACL, 2020. Acceptance rate: 22%. [\[link\]](#)
- CSCW 2020b Configuring Audiences: A Case Study of Email Communication
Justine Zhang, James W. Pennebaker, Susan T. Dumais, and Eric Horvitz.
Proceedings of CSCW, 2020. [\[link\]](#)
- SIGDIAL 2020 ConvoKit: A Toolkit for the Analysis of Conversations
Jonathan P. Chang, Caleb Chiam, Liye Fu, Andrew Z. Wang, **Justine Zhang**,
and Cristian Danescu-Niculescu-Mizil.
Proceedings of SIGDIAL, 2020. System demo. [\[link\]](#)

- ACL 2019 Finding Your Voice: The Linguistic Development of Mental Health Counselors
Justine Zhang, Robert Filbin, Christine Morrison, Jaclyn Weiser,
 and Cristian Danescu-Niculescu-Mizil.
 Proceedings of ACL, 2019. Acceptance rate: 23%. [\[link\]](#)
- CSCW 2018 Characterizing online public discussions through patterns of participant interactions
Justine Zhang, Cristian Danescu-Niculescu-Mizil, Christina Sauper,
 and Sean J. Taylor
 Proceedings of CSCW, 2018. Acceptance rate: 26%. [\[link\]](#)
- ACL 2018 Conversations Gone Awry: Detecting Early Signs of Conversational Failure
Justine Zhang, Jonathan P. Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon,
 Yiqing Hua, Dario Taraborelli, and Nithum Thain
 Proceedings of ACL, 2018. Acceptance rate: 25%. [\[link\]](#)
- EMNLP 2017 Asking too much? The rhetorical role of questions in political discourse.
Justine Zhang, Arthur Spirling, and Cristian Danescu-Niculescu-Mizil.
 Proceedings of EMNLP, 2017. Acceptance rate: 26%. [\[link\]](#)
- ICWSM 2017a Community Identity and User Engagement in a Multi-Community Landscape.
Justine Zhang*, William L. Hamilton*, Cristian Danescu-Niculescu-Mizil,
 Dan Jurafsky, and Jure Leskovec.
 Proceedings of ICWSM, 2017. Acceptance rate: 14%. [\[link\]](#)
- ICWSM 2017b Loyalty in Online Communities.
 William L. Hamilton*, **Justine Zhang***, Cristian Danescu-Niculescu-Mizil,
 Dan Jurafsky, and Jure Leskovec.
 Proceedings of ICWSM, 2017. Short paper. [\[link\]](#)
- NAACL 2016 Conversation flow in Oxford-style debates.
Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil.
 Proceedings of NAACL, 2016. Short paper. Acceptance rate: 29%. [\[link\]](#)
- WWW 2015 QUOTUS: The Structure of Political Media Coverage as Revealed by Quoting
 Patterns.
 Vlad Niculae*, Caroline Suen*, **Justine Zhang***, Cristian Danescu-Niculescu-Mizil,
 and Jure Leskovec.
 Proceedings of WWW, 2015. Acceptance rate: 14%. [\[link\]](#)

External Research Experience

- | | |
|--|------|
| Research Intern , <i>Microsoft Research</i> | 2019 |
| Worked with Susan Dumais, Eric Horvitz, and James W. Pennebaker
Examining audience configurations in emailing | |
| Research Fellow , <i>Crisis Text Line</i> | 2018 |
| In collaboration with Cristian Danescu-Niculescu-Mizil, analyzing crisis
counseling conversations; developing models of conversational dynamics | |
| Research Intern , <i>Facebook</i> | 2017 |
| Core Data Science team, worked with Sean J. Taylor and Christina Sauper
Characterizing public conversations | |

Invited Talks

- Towards an Actionable Understanding of Conversations** 2021
Emory QTM, EPFL IC, Northeastern CS, UChicago CS, WUSTL CS, MPI SWS, CMU LTI, UMich School of Information
- Interdisciplinary Seminar in Quantitative Methods, University of Michigan** [\[link\]](#) 2019
Presented research on modeling conversations and causal inference in conversations
- Text as Data Speaker Series, New York University** [\[link\]](#) 2018
Presented research on modeling conversation dynamics

Invited Workshop Participation

- EECS Rising Stars workshop, UC Berkeley** 2020
- Microsoft Research AI Breakthroughs workshop** 2019, 2020
Presented posters on unsupervised modeling of conversation dynamics
- Text as Data workshop, Stanford University** 2019
Presented research on causal inference in conversations
- Microsoft Research PhD Summit** 2018
Presented poster on modeling the rhetorical intentions of questions
- Google NLP/NLU workshop** 2017
26 faculty and 20 students invited
Presented poster on conversation dynamics in public debates
- Text as Data workshop, Northeastern University** 2016
Presented research on question dodging in debates

Media Coverage

- Cornell Chronicle:** AI analysis gives guidance to crisis counselors [\[link\]](#)
- Fast Company:** Scientists are building a detector for conversations likely to go bad [\[link\]](#)
- MIT Technology Review:** Machine learning could stop an online war of words before it starts [\[link\]](#)
- The Verge:** Machine learning is helping computers spot arguments online before they happen [\[link\]](#)

Service

- Area chair:** NAACL 2021, Computational Social Science and Social Media track
- Reviewer (conferences):** ACL (2018, 2019, 2020); CSCW (2018, 2020, 2021); CHI (2020); EMNLP (2017, 2018, 2019); ICWSM (2018, 2019, 2020, 2021); NAACL (2019)
- Reviewer (journals):** TKDD; First Monday; PLOS ONE
- Reviewer (workshops):** ACL SRW (2019); EMNLP WNUT (2018); NAACL SRW (2018); WebSci Poster Track (2019), EACL HCI+NLP Workshop (2021)

Data and Code

ConvoKit [\[link\]](#): Open-source toolkit to analyze conversational behaviours.
Includes implementations of methodology and analyses for the following publications:
EMNLP 2017, ACL 2018, CSCW 2018, ACL 2019, ACL 2020 and SIGDIAL 2020 papers.

Datasets released via ConvoKit: Parliamentary questions period corpus (EMNLP 2017),
Conversations Gone Awry dataset (ACL 2018), Supreme Court Oral Arguments corpus (ACL 2020)

QUOTUS [\[link\]](#): Interactive visualization of media coverage of political speeches, and dataset, for
WWW 2015 paper.

Intelligence Squared Debates Dataset [\[link\]](#): For NAACL 2016 paper

References

Cristian Danescu-Niculescu-Mizil

Associate Professor, Department of Information Science
Cornell University
cristian@cs.cornell.edu

Susan Dumais

Technical Fellow & Managing Director
Microsoft Research New England, New York City and Montreal
sdumais@microsoft.com

Eric Horvitz

Technical Fellow and Chief Scientific Officer
Microsoft
horvitz@microsoft.com

Dan Jurafsky

Jackson Eli Reynolds Professor in Humanities, Professor of Linguistics, Professor of Computer
Science
Stanford University
jurafsky@stanford.edu

Sendhil Mullainathan

Roman Family University Professor of Computation and Behavioral Science
The University of Chicago, Booth School of Business
sendhil@chicagobooth.edu