

Junyi Jessy Li

Assistant Professor

Department of Linguistics
The University of Texas at Austin
✉ jessy@austin.utexas.edu
📧 jessyli.com

Research Interests

Computational Linguistics and Natural Language Processing with a focus on discourse and pragmatics.

Education

- 2012-2017 **Ph.D. in Computer and Information Science**, University of Pennsylvania.
Thesis: From Discourse Structure to Text Specificity: Studies of Coherence Preferences
Advisor: Prof. Ani Nenkova
- 2011-2013 **M.Sc. in Engineering, Computer and Information Science**, University of Pennsylvania.
- 2009-2011 **B.Sc. in Engineering, Computer Science and Engineering**, University of Michigan—Ann Arbor.
- 2007-2011 **B. in Engineering, Electrical and Computer Engineering**, Shanghai Jiao Tong University, China.

Referred Publications

- [1] Wei-Jen Ko, Greg Durrett, and **Junyi Jessy Li**. Domain agnostic real-valued specificity prediction. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019. to appear.
- [2] Yifan Gao, Yang Zhong, Daniel Preoțiuc-Pietro, and **Junyi Jessy Li**. Predicting and analyzing language specificity in social media posts. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2019. to appear.
- [3] Eric Holgate, Isabel Cachola, Daniel Preoțiuc-Pietro, and **Junyi Jessy Li**. Why swear? Analyzing and inferring the intentions of vulgar expressions. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4405–4414, 2018.
- [4] Isabel Cachola, Eric Holgate, Daniel Preoțiuc-Pietro, and **Junyi Jessy Li**. Expressively vulgar: The socio-dynamics of vulgarity and its effects on sentiment analysis in social media. In *International Conference on Computational Linguistics (COLING)*, pages 2927–2938, 2018. This paper was named an **Area Chair Favorite**.
- [5] Benjamin Nye, **Junyi Jessy Li**, Roma Patel, Yinfei Yang, Iain Marshall, Ani Nenkova, and Byron Wallace. A corpus with multi-level annotations of patients, interventions and outcomes to support language processing for medical literature. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 197–207, 2018.
- [6] Pengyu Nie, **Junyi Jessy Li**, Sarfraz Khurshid, Raymond Mooney, and Milos Gligoric. Natural language processing and program analysis for supporting todo comments as software evolves. In *AAAI Workshop on NLP for Software Engineering (NL4SE)*, page 775–778, 2018.
- [7] An Thanh Nguyen, Byron Wallace, **Junyi Jessy Li**, Ani Nenkova, and Matthew Lease. Aggregating and predicting sequence labels from crowd annotations. In *Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 299–309, 2017.
- [8] **Junyi Jessy Li**, Julia Parish-Morris, Leila Bateman, and Ani Nenkova. Autism quotient scores modulate the perception and production of text specificity in adult females. In *International Meeting for Autism Research (IMFAR)*, 2017. Abstract.

- [9] **Junyi Jessie Li**, Kapil Thadani, and Amanda Stent. The role of discourse units in near-extractive summarization. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 137–147, 2016. This paper was **nominated for the Best Paper Award**.
- [10] **Junyi Jessie Li** and Ani Nenkova. The instantiation discourse relation: A corpus analysis of its properties and improved detection. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, short papers (NAACL)*, pages 1181–1186, 2016.
- [11] **Junyi Jessie Li**, Bridget O’Daniel, Yi Wu, Wenli Zhao, and Ani Nenkova. Improving the annotation of sentence specificity. In *International Conference on Language Resources and Evaluation (LREC)*, pages 3921–3927, 2016.
- [12] **Junyi Jessie Li**. Estimating text intelligibility via information packaging analysis. In *AAAI Conference on Artificial Intelligence: Doctoral Consortium (AAAI-DC)*, pages 4305–4306, 2016.
- [13] **Junyi Jessie Li** and Ani Nenkova. Detecting content-heavy sentences: A cross-language case study. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1271–1281, 2015.
- [14] **Junyi Jessie Li** and Ani Nenkova. Fast and accurate prediction of sentence specificity. In *AAAI Conference on Artificial Intelligence (AAAI)*, pages 2281–2287, 2015.
- [15] **Junyi Jessie Li**, Marine Carpuat, and Ani Nenkova. Cross-lingual discourse relation analysis: A corpus study and a semi-supervised classification system. In *International Conference on Computational Linguistics (COLING)*, pages 577–587, 2014.
- [16] **Junyi Jessie Li**, Marine Carpuat, and Ani Nenkova. Assessing the discourse factors that influence the quality of machine translation. In *Annual Meeting of the Association for Computational Linguistics, short papers (ACL)*, pages 283–288, 2014.
- [17] **Junyi Jessie Li** and Ani Nenkova. Addressing class imbalance for improved recognition of implicit discourse relations. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 142–150, 2014.
- [18] **Junyi Jessie Li** and Ani Nenkova. Reducing sparsity improves the recognition of implicit discourse relations. In *Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 199–207, 2014.

Patents

Lucretia H. Vanderwende, Sumit Basu, Charles E. Jacobs and **Junyi Li**. Generating High-Level Questions from Sentences. *US Patent Application US 20160063879 A1 [pending]*.

Professional Experience

- Since Fall 2017 **Assistant Professor**, *Department of Linguistics, The University of Texas at Austin*.
- Summer 2015 **Research Intern**, *Yahoo! Labs*, New York, NY.
Worked with Amanda Stent and Kapil Thadani on automatic summarization using discourse units, rhetorical structure and local coherence. Published paper [2].
- Summer 2013 **Research Intern**, *Microsoft Research*, Redmond, WA.
Worked with Lucy Vanderwende and Sumit Basu on high level question generation in unstructured text using discourse analysis. Submitted patent.
- Summer 2010 **Research Assistant**, *University of Michigan*, Ann Arbor, MI.
Worked with Eytan Adar on context-driven image retrieval and structured data visualization.

Teaching

- Fall 2018 **Instructor**, *Language and Computers (LIN 313)*, The University of Texas at Austin.
- Fall 2018 **Instructor**, *Computational Discourse (LIN 393S)*, The University of Texas at Austin.
- Spring 2018 **Instructor**, *Language and Computers (LIN 313)*, The University of Texas at Austin.
- Spring 2018 **Instructor**, *Research in Computational Linguistics (LIN 389C)*, The University of Texas at Austin.
- Fall 2014 **Head Teaching Assistant**, *Computational Linguistics (CIS 530)*, University of Pennsylvania.
- Fall 2013 **Teaching Assistant**, *Computational Linguistics (CIS 530)*, University of Pennsylvania.
- Fall 2012 **Teaching Assistant**, *Computational Linguistics (CIS 530)*, University of Pennsylvania.
- Spring 2012 **Teaching Assistant**, *Programming Languages and Techniques I (CIS 120)*, University of Pennsylvania.
- Spring 2009 **Teaching Assistant**, *Introduction to Programming (VG 101)*, Shanghai Jiao Tong University.

Professional Activities

Committee Service

- NAACL 2019 **PC Member**, North American Chapter of the Association for Computational Linguistics — Human Language Technologies.
- AAAI 2019 **PC Member**, AAAI Conference on Artificial Intelligence.
- NAACL 2018 **Area Chair**, *Discourse and Pragmatics*, North American Chapter of the Association for Computational Linguistics — Human Language Technologies.
- ACL 2018 **PC Member**, The Association of Computational Linguistics.
- EMNLP 2018 **PC Member**, Empirical Methods in Natural Language Processing.
- AAAI 2018 **PC Member**, AAAI Conference on Artificial Intelligence.
- IJCAI 2018 **PC Member**, International Joint Conference on Artificial Intelligence.
- COLING 2018 **PC Member**, International Conference on Computational Linguistics.
- SIGDIAL 2018 **PC Member**, Special Interest Group in Discourse and Dialogue.
- WiNLP 2018 **PC Member**, Widening Natural Language Processing Workshop.
- W-NUT 2018 **PC Member**, Workshop on Noisy User-generated Text.
- NLPCC 2018 **PC Member**, International Conference on Natural Language Processing and Chinese Computing.
- EMNLP 2017 **PC Member**, Empirical Methods in Natural Language Processing.
- ACL 2017 **PC Member**, The Association of Computational Linguistics.
- ACL-SRW 2017 **PC Member**, Student Research Workshop at the Association of Computational Linguistics.
- EACL 2017 **PC Member**, European Chapter of the Association of Computational Linguistics.
- EACL-SRW 2017 **PC Member**, Student Research Workshop at the European Chapter of the Association of Computational Linguistics.
- W-NUT 2017 **PC Member**, Workshop on Noisy User-generated Text.
- NewSum 2017 **PC Member**, Workshop on New Frontiers in Summarization.
- NAACL 2016 **PC Member**, North American Chapter of the Association for Computational Linguistics — Human Language Technologies. (**Named among Best Reviewers**).
- EMNLP 2016 **PC Member**, Conference on Empirical Methods in Natural Language Processing.
- COLING 2016 **PC Member**, International Conference on Computational Linguistics.

- ACL-SRW 2016 **PC Member and Student Mentor**, Student Research Workshop at the Association of Computational Linguistics.
- W-NUT 2016 **PC Member**, Workshop on Noisy User-generated Text.
- MASC 2016 **Session Chair and Volunteer**, Mid-Atlantic Student Colloquium on Speech, Language and Learning.
- NAACL 2015 **PC Member**, North American Chapter of the Association for Computational Linguistics — Human Language Technologies.
- CoNLL 2015 **PC Member**, The Conference on Natural Language Learning Shared Task: Shallow Discourse Parsing.
- MASC 2015 **PC Member**, Mid-Atlantic Student Colloquium on Speech, Language and Learning.
- YRRSDS 2014 **Local Organizer**, Young Researchers' Roundtable on Spoken Dialog Systems.
- ACL-SRW 2013 **PC Member**, Student Research Workshop at the Association of Computational Linguistics.
- Reviewer**
- 2018 Dialogue and Discourse (Journal).
- 2018 Computational Linguistics (Journal).
- 2017 Dialogue and Discourse (Journal).
- WWW 2015 International World Wide Web Conference.
- AAAI 2015 AAAI Conference on Artificial Intelligence.
- RANLP 2015 Recent Advances in Natural Language Processing.
- COLING 2014 International Conference on Computational Linguistics.
- RANLP 2013 Recent Advances in Natural Language Processing.
- Seminar Organizer**
- 2013-2015 Computational Linguistics Colloquium Series, University of Pennsylvania.

Grants and Awards

- 2018 **Salesforce Research Deep Learning Grant**, “*Hierarchical Graph-based Advice Summarization from Online Forums*”, co-PI: Prof. Katrin Erk, \$50K.
- 2018 **Amazon Alexa Fund**, “*Generating Informative Conversation Responses with Text Specificity and Opportunistic Learning*”, co-PI: Asst. Prof. Greg Durrett, Prof. Raymond J. Mooney, \$63K.
- 2018 **Amazon Web Services Cloud Credits for Research**, with Asst. Prof. Greg Durrett, Prof. Raymond J. Mooney, \$13.2K AWS credits.
- 2018 **Summer Research Assignment**, Dean of Graduate Studies, The University of Texas at Austin, two months summer salary.
- 2018 **Area Chair Favorite**, International Conference on Computational Linguistics (COLING 2018).
- 2016 **Best Paper Nominee**, Special Interest Group on Discourse and Dialogue (SIGDIAL 2016).
- 2016 **Best Reviewer**, North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL 2016).
- 2016 **Student Scholarship**, AAAI/SIGAI Doctoral Consortium (AAAI 2016).
- 2011 **Magna Cum Laude**, University of Michigan.
- 2010, 2011 **University Honors**, University of Michigan.
- 2009-2011 **Dean's List**, University of Michigan.
- 2009 **Outstanding Scholarship**, Shanghai Jiao Tong University.
- 2007-2009 **Dean's List**, Shanghai Jiao Tong University.

Invited Talks

Assessing Text Intelligibility with Discourse Structure

- 03/2017 Texas State University
- 03/2017 University of California, Davis
- 02/2017 Pennsylvania State University
- 02/2017 The University of Texas at Austin, Department of Linguistics
- 02/2017 University of Illinois at Chicago
- 02/2017 University of Illinois at Urbana-Champaign
- 02/2017 Washington University in St. Louis
- 02/2017 University of Iowa
- 01/2017 The University of Texas at Austin, School of Information

Text Specificity: How and Why

- 11/2016 Bloomberg LP
- 10/2016 The University of Texas at Austin, Department of Computer Science

Spiceteller—an NLP tool developed at UPenn

- 06/2016 Philadelphia Python Users Group

Understanding Text Specificity

- 06/2016 Interactions Corporation

How does the computer understand language?

- 04/2016 Women in Computer Science High School Open Day, University of Pennsylvania
- 03/2016 Lower Merion High School

Analyzing General and Specific Content in Text

- 11/2015 University of Maryland, College Park

Predicting Sentence Specificity

- 09/2015 Center for Undergraduate Research and Fellowships, University of Pennsylvania

Content-heavy Sentences in Cross-lingual Context

- 09/2015 Columbia University