

# Lucy Li

Natural language processing, computational sociolinguistics, & computational social science

✉ lucy3\_li@berkeley.edu | 🏠 lucy3.github.io

## Education

---

### University of California, Berkeley

PhD Information Science

- Advisor: David Bamman

Berkeley, CA

Aug 2019 - present

### Stanford University

BS Symbolic Systems, MS Computer Science

- A coterminal degree, w/ a concentration in language and depth in artificial intelligence.
- Study abroad at University of Oxford, Winter 2017.

Stanford, CA

Sept 2014 - June 2019

## Experience

---

### Microsoft Research

Research Intern

- Mentors: Alexandra Olteanu, Su Lin Blodgett
- Fairness, Accountability, Transparency, Ethics (FATE) team.

Montreal, Canada

May 2021 - Aug 2021

### Stanford Computer Science

Research Assistant

- Advisors: Dan Jurafsky, Patricia Bromley.
- Investigated the framing and representation of underrepresented groups in history textbooks with linguistics PhD student Dora Demszky.

Stanford, CA

Jan 2019 - Dec 2019

### École Polytechnique Fédérale de Lausanne

Research Intern

- Advisor: Robert West (Data Science Lab)
- Operationalized and analyzed behavioral trends in a political quote dataset using Apache Spark, emotion lexicons, Stanford CoreNLP parsers, and social networks.

Lausanne, Switzerland

July 2018 - Sept 2018

### Stanford Computer Science

Research Assistant

- Advisors: David Jurgens, Jure Leskovec (Stanford Network Analysis Project), Dan Jurafsky (NLP group)
- Used language and social network features to classify fictional and real relationships with scikit-learn, NLTK, and Keras.

Stanford, CA

April 2017 - June 2018

## Publications

---

\*indicates equal contribution

**Li Lucy**, David Bamman. Gender and Representation Bias in GPT-3 Generated Stories. *Workshop on Narrative Understanding (WNU) at the North American Association for Computational Linguistics (NAACL)*, 2021.

**Li Lucy**, David Bamman. Characterizing English variation across social media communities with BERT. *Transactions of the Association of Computational Linguistics (TACL)*, 2021.

**Li Lucy\***, Dora Demszky\*, Patricia Bromley, Dan Jurafsky. Content Analysis of Textbooks via Natural Language Processing: Findings on Gender, Race, and Ethnicity in Texas U.S. History Textbooks. *AERA Open*, 2020. [Best paper at American Educational Research Association (AERA) Educational Data Science Conference.]

Emma Lurie, **Li Lucy**, Masha Belyi, Sofia Dewar, Daniel Rincón, John Baldwin, Rajvardhan Oak. Investigating Causal Effects of Instructions in Crowdsourced Claim Matching. *Computation + Journalism Symposium (C+J)*, 2020. [non-archival.]

**Li Lucy**, Julia Mendelsohn. Using sentiment induction to understand variation in gendered online communities. *Society for Computation in Linguistics (SCiL)*, 2019.

**Li Lucy**, Jon Gauthier. Are distributional representations ready for the real world? Evaluating word vectors for grounded perceptual meaning. *Language Grounding for Robotics (RoboNLP) Workshop at the Association for Computational Linguistics (ACL)*, 2017.

## Awards, Fellowships, & Grants

---

<b>Graduate Research Fellowship</b> , National Science Foundation	2019
<b>K. Jon Barwise Award for Distinguished Contributions</b> , Stanford Symbolic Systems	2018
<b>Undergraduate Advising &amp; Research (UAR) Small Grant</b> , \$1500, Stanford University	2018
<b>Grants for Education and Research</b> , \$1145, Stanford Symbolic Systems	2017
<b>Phi Beta Kappa</b> , Stanford University	2017

## Presentations

---

*Characterizing English variation across social media communities with BERT.*  
June 2021. Guest lecture, “Computational Text Analysis,” Barnard College.

*Content Analysis of Textbooks via Natural Language Processing.*  
Feb 2021. Guest lecture, “Doing Digital History,” Stanford.  
Feb 2021. Stanford Human-Computer Interaction Lunch Seminar.  
May 2021. Guest lecture, “Using Data to Describe the World,” Stanford.  
May 2020. Guest lecture, “Using Data to Describe the World,” Stanford.  
Oct 2019. 10th Annual New Directions in Analyzing Text as Data (TADA). Stanford, CA.

## Teaching Experience

---

<b>Stanford CS 224U</b> , Course Assistant, top 5% in computer science	Spring 2019
<b>Symbolic Systems Program</b> , Advising Fellow	2016 - 2017, 2019
<b>Stanford EE/CME 103</b> , Course Assistant	Fall 2017

## Service

---

### *Professional*

**Reviewer:** ACL (2021), ACL Workshop on NLP for Positive Impact (2021), EMNLP (2021), AERA Open (2020), NeurIPS Human and Machine in-the-Loop Evaluation and Learning Strategies Workshop (2020).

**Organizing committee:** Teaching NLP (2021) at NAACL.

### *Community*

**General:** Diaries of Social Data Research (Podcast Host), Sociologists of Digital Things (Admin).

**Students:** BAIR Mentoring Program (2020-Present), CS Kickstart (Speaker; 2020), UC Berkeley Girls in Engineering (Leader; 2020), Berkeley AI4ALL (Mentor; 2019), Stanford AI4All (Mentor; 2019), Girls Teaching Girls to Code (Mentor/Lead; 2018, 2019).

## Skills

---

**Computer Languages:** Python, Julia, C++, SQL

**Natural Languages:** English, Mandarin Chinese

**Tools:** NLTK, Stanford CoreNLP, SpaCy, scikit-learn, Apache Spark, MTurk, Figure Eight, Keras, TensorFlow, PyTorch.