

HANG JIANG

47 Olmsted Rd, Escondido Village, Studio 6 148T, Stanford, California 94305
(470)-270-7700; hjian42@stanford.edu
Homepage: <https://hjian42.github.io/>

OBJECTIVE I am actively applying for Ph.D programs in natural language processing, computational psycholinguistics, and computational social science for Fall 2020.

EDUCATION **Stanford University**, Palo Alto, CA Sept. 2018 - Exp. Jun. 2020
M.S. in Symbolic Systems, *focus on AI and NLP* GPA: 4.03/4.00
Advisor: Michael C. Frank

Emory University, Atlanta, GA Aug. 2014 - May. 2018
B.S. in Computer Science and B.A. in Linguistics, *Summa Cum Laude*
Advisor: Jinho D. Choi
Thesis: Automatic Personality Prediction with Attention-based Neural Networks

RESEARCH INTEREST

- Machine Learning, Natural Language Processing, Computational Linguistics
- Computational Social Science, Computational Cognitive Science
- Language Variation and Change, Social Reasoning, Language Acquisition
- Machine Learning for Education and Sustainable Development

PUBLICATIONS **Jiang, Hang**; Xianzhe Zhang; and Jinho D. Choi. 2019. Automatic Text-based Personality Recognition on Monologues and Multiparty Dialogues Using Attentive Networks and Contextual Embeddings. To appear in *Proceedings of the 34th AAAI Conference on Artificial Intelligence: Student Abstract and Poster Program, of AAAI:SAP'19*. New York, USA. [[paper](#), [github](#), [data](#)] **Strong Acceptance.**

Jiang, Hang*; Haoshen Hong*; Yuxing Chen*; and Vivek Kulkarni. 2019. Dialect-Gram: Automatic Detection of Dialectal Changes with Multi-geographic Resolution Analysis. To appear in *Proceedings of the Society for Computation in Linguistics. New Orleans: Linguistic Society of America*. **Oral Presentation.** [[paper](#), [github](#), [data](#)]

Jiang, Hang; and Jinho D. Choi. 2018. Automatic Personality Prediction with Attention-based Neural Networks. *Electronic Theses and Dissertations*, Emory University, Atlanta, GA. **Highest Honors.** [[paper](#), [github](#)]

TALKS & PRESENTATIONS **Jiang, Hang**; and Roberto Franzosi. 2018. Shape of Stories. *Emory SIAM Student Chapter Event "Bridging the Gap: Math across Emory"*, Atlanta, GA. [[slides](#)]

Jiang, Hang; Doris Zhou; Alec Wolyniec; and Roberto Franzosi. 2017. Automatic Extraction of Actors and Actions from News Stories of Lynching. *Mini-Conference: Lynching in Historical Perspective*, Atlanta, GA. [[slides](#)]

Zhou, Doris; **Hang Jiang**; and Roberto Franzosi. 2017. He, She, They... Who Are They? Automatic and Semi-Automatic Anaphora Resolution. *Mini-Conference: Lynching in Historical Perspective*, Atlanta, GA. [[slides](#)]

Xiao, Catherine; **Hang Jiang**; and Jinho D. Choi. 2017. Seq2seq Model to Tokenize the Chinese Language. *Emory Undergraduate Research Symposium in Fall, Atlanta, GA*. [poster]

Jiang, Hang; and Jinho D. Choi. 2016. Chinese Grammar vs English Grammar in Universal Dependency. *Emory Undergraduate Research Symposium in Spring, Atlanta, GA*. [poster]

INDUSTRY EXPERIENCE

Apple Inc., Cupertino, California Jun. 2019 - Sept. 2019
Machine Learning Applied to Natural Language Processing Intern

Supervisors: Bing Zhao, Vivek Kumar Rangarajan Sridhar

- Worked on sentence embeddings, applied in text summarization and smart reply
- Contributed to internal NLP framework, supporting Apple Core ML

Educational Testing Service (ETS), Princeton, NJ Jun. 2018 - Aug. 2018
Natural Language Processing Intern

Supervisors: Martin Chodorow, Nitin Madnani, Aoife Cahill

- Built a new automatic *preposition error detection* system for *E-rater*
- Improved recall by **10%** to help ETS in grading GRE and TOEFL essays

RESEARCH EXPERIENCE

Stanford Language and Cognition Lab Mar. 2019 - Exp. Jun. 2020
Research Assistant Supervisors: Abdellah Fourtassi, Vivek Kulkarni, Michael C. Frank

- design both semantic networks and apply diachronic word embeddings to study semantic development during child language learning on CHILDES database
- Working on a paper towards *Association of Computational Linguistics 2020*

Stanford University Jan. 2019 - Exp. Jan. 2020
Independent Research

- Published *DialectGram: Automatic Detection of Dialectal Changes with Multi-geographic Resolution Analysis* with Vivek Kulkarni (Stanford NLP Lab) at Society for Computation in Linguistics (SCiL 2020)
- Working on a project on automatic earthquake detection with Weiqiang Zhu (Geophysics Lab) and Kai Sheng Tai (Stanford Future Data Systems Group)

Honors Thesis Program, Emory NLP Lab Sept. 2017 - May. 2018
Research Assistant Supervisor: Jinho D. Choi

Committee Members: Marjorie Pak, Roberto Franzosi, Shun Yan Cheung

- Constructed FriendsPersona, the first dialogue dataset for automatic personality recognition from Friends TV show transcripts with Amazon's Mechanical Turk
- Improved state-of-art performance on Essays dataset with attentive networks and contextual embeddings and established a strong benchmark on FriendsPersona
- Won *highest honors* and submitted a paper to AAAI-20 Student Abstract and Poster Program

Emory PC-ACE Lab Jan. 2017 - May. 2018
Research Assistant Supervisor: Roberto Franzosi

- Led the development of Emory Program for Computer-Assisted Coding of Events (PC-ACE) key components including automatic subject-verb-object (SVO) extraction, sentiment analysis, co-reference resolution, document clustering, and data visualization for narrative analysis (*Java, Python*)

- Assisted students in *LING 446W: Big/Small Data and Visualization* to use PC-ACE software for research in computational social science (*NLTK*, *Scikit-Learn*, *Scipy*, *Stanford CoreNLP*)

Carnegie Mellon University Jun. 2017 - Aug. 2017

Summer Research Intern Supervisor: Brian MacWhinney

- Built and optimized automatic speech recognition (ASR) models based on *Bidirectional LSTM* to recognize speech from Aphasia patients
- Developed an intelligent English tutor *Funetics* in *Nodejs*, *Express*, *MongoDB* and *jQuery* to grade speech and provide customized feedback

Georgia Institute of Technology Mar. 2017 - May. 2017

Research Assistant Supervisor: Annamaria Conti

- Implemented *Latent Semantic Analysis (LSA)* and *Latent Dirichlet Allocation (LDA)* to find patent-paper pairs on a large biomedical dataset
- Identified 688 high-quality patent-paper pairs and was acknowledged for model implementation

Summer Research Program, Emory NLP Lab Jun. 2016 - Sept. 2016

Summer Research Intern Supervisor: Jinho D. Choi

- Built a Seq2Seq-based *Chinese Word Segmentation* model on Chinese TreeBank 7.0 in *Tensorflow*; Presented the work at Emory Fall Symposium

PROJECTS

- Ensemble BERT with Data Augmentation and Linguistic Knowledge on SQuAD 2.0, CS224N Deep Learning for Natural Language Processing, 2019, Christopher Manning. **Ranked #1 on the leaderboard and won the best poster award for the default project division (1 out of 98 teams)** [[report](#), [poster](#), [twitter](#)]
- Improving Sentiment Analysis with Data Augmentation, CS229A Applied Machine Learning, Younes Bensouda Mourri, Andrew Ng, 2019, [[report](#), [poster](#)]
- JoyBot: Your Personal Mood Tracker, Apple iContest for Interns. [[demo](#)]

COURSES

Stanford University

- **CS:** Natural Language Processing in Deep Learning (Chris Manning), Information Retrieval (Chris Manning), Natural Language Understanding (Chris Potts), Programming for Linguists (Chris Potts), From Languages to Information (Dan Jurafsky), Applied Machine Learning (Andrew Ng), Decision Making Under Uncertainty (Mykel Kochenderfer), Data Management and Data Systems (Shiva Shivakumar), Object Oriented System Design (Patrick Young), Programming Abstractions in C++ (Marty Stepp)
- **Planned:** Mining Massive Data Sets (Jurec Leskovec), Principles of Computer Systems (Jerry Cain), Probabilistic Graphical Models (Stefano Ermon), Convolutional Neural Networks for Visual Recognition (Feifei Li)

Emory University

- **CS, Math:** Object-Oriented Programming in Java, Data Structure and Algorithms, Database Systems, Computational Linguistics, Data Mining, System Programming, Theory of Computing, Computer Networks, Discrete Structures, Linear Algebra, Optimization Theory, Statistics Inference, Foundations of Math, Big/Small Data & Visualization
- **Linguistics, Psychology, Philosophy:** Language Mind and Society, Semantics and Pragmatics, Phonetics and Phonology, Syntax and Morphology, Brain

and Language, History of the American Languages, Foundations of Linguistics, Cognition, Logic, Psychology I & II

| | | |
|------------------------------------|--|-----------|
| EXTRA-CURRICULAR ACTIVITIES | Activity Manager, Chinese Entrepreneurship Organization (CEO) | 2018-2020 |
| | Member, Symbolic Systems Society at Stanford | 2018-2020 |
| | Member, Association of Chinese Students and Scholars at Stanford | 2018-2020 |
| | Vice President in Public Management, Emory Linguistics Circle | 2017 |
| | Campus Leader, Emory Bread House | 2016-2017 |
| | Member, Structured Independent Language Study (SILS) Program | 2016-2017 |
| | Vice President in Management, Sino-Emory Newsletter | 2017 |
| | Editor, Sino-Emory Newsletter | 2016 |
| | Member, Artisan Guild at Emory | 2016 |
| HONORS | Grants for Education And Research (GEAR) , Stanford Research Funding | 2019 |
| | Best Poster , CS224n NLP with Deep Learning Final Project (Chris Manning) | 2019 |
| | Highest Honors , Honors Thesis Program in Linguistics at Emory University | 2018 |
| | 2nd Place , CS378 Data Mining Algorithm Competition at Emory University | 2017 |
| | Phi Sigma Iota Honor Society , International Foreign Language Studies | 2017 |
| | Dean List , Emory College of Arts & Sciences | 2017 |
| | Research Fellowship , Scholarly Inquiry and Research at Emory Program | 2016 |
| TEACHING | Research Fellowship , Scholarly Inquiry and Research at Emory Program | 2015 |
| | Teaching Assistant: CS 145 Data Management and Data Systems, Stanford University (Autumn 2019) Instructor: Shiva Shivakumar | |
| | | |
| SKILLS | Programming Languages: Python, C/C++, Java, SQL, R, Javascript, Swift, HTML/CSS, M68000 Assembly | |
| | Frameworks & Tools: Tensorflow, PyTorch, Android, iOS, JQuery, Git, LaTeX | |
| | Natural Languages: Mandarin Chinese (native), English (fluent), Portuguese (intermediate), Spanish (intermediate), American Sign Language (beginner), Korean (beginner), Vietnamese (beginner) | |
| | Certificates: Natural Language Processing Nanodegree (Udacity), Deep Learning Specialization (Coursera) | |
| | | |
| | | |
| | | |