Siyuan Song

Candidate for B.A. in Linguistics, Shanghai Jiao Tong University
sjtusongsy2022@sjtu.edu.cn — +86 18951948778 — 800 Dongchuan Road, Shanghai, China — SiyuanSong2004.github.io

RESEARCH INTERESTS

I am interested in abstracting the meaning of complex languages using computational methods and building models for analyzing or simulating human cognition, especially in language processing. My previous work has focused on exploring the shortcomings of existing models when dealing with complex languages phenomenons, and advising computer scientists and technology companies on linguistic perspectives.

Recently, I have been impressed by Geoffrey Hinton's questioning of the Chomskyan idea of innate competence. I hope to further reveal the mechanisms behind how machines and humans process language through future research, (maybe) defending the viewpoints of past linguists.

Therefore, I'm especially interested in **Computational Linguistics, Psycholinguistics and Cognitive Science**.

EDUCATION

Shanghai Jiao Tong University, Shanghai, China   Aug 2022 — Present
Candidate for B.A. in Linguistics

Nanjing Foreign Language School, Nanjing, China   Sep 2016 — June 2022

RESEARCH PROJECTS

Testing Language Models on Chinese Xiehouyu Tasks   Shanghai Shi, China
Project Leader, supervised by Prof. Hai Hu   September 2023 — Current

- Initiated a personal research project focusing on Chinese Xiehouyu (a type of language riddle)
- Designed tasks based on Xiehouyu to challenge and analyze the inferential understanding capabilities of LLMs
- Tried to answer following research questions: Can language models understand riddles with figurative language, allusions or homonyms? Do they rely on context or existing data?
- Gained hands-on experience in preprocessing language data and testing Generative LMs
- Developed foundational skills in language data handling and coding

Collaborators: Kejia Zhang

Evaluating LLM's Ability in Pragmatics: the SwordImp Dataset   Shanghai Shi, China
Key Member, supervised by Prof. Hai Hu   September 2023 — Current

- Assisted in the development of the SwordImp dataset, aimed at testing large language models' understanding of Sentence Implicature
- Curated sentences based on the Gricean Maxim from original texts and crafted Distractors following specific rules
- Helped in collecting data for human baseline experiments

Collaborators: Shisen Yue (Project Leader), Xinyuan Cheng

Exploring large models' understanding of Binding Principles   Shanghai Shi, China
Key Member, supervised by Prof. Hai Hu   February 2024 — Current

- Designed Minimal Pairs to test model mastery of the binding principle
- Tested whether in-context learning could enhance model robustness in syntax performance
- This research is still in its early stages, and more information will be updated.

Collaborators: Yikang Liu, Dr. Lilong Xu, Hongao Zhu

Student Research Project on Language Models and Chinese Idioms   Shanghai Shi, China
Student Researcher   September 2022 — March 2024

- Conducted in-depth data collection from dictionaries and corpora, annotated sentences for precise model training, and performed extensive tests to assess model performance.
- Utilized a RoBERTa-based NLI model for testing, employing Python modules like numpy for comprehensive data analysis, enhancing the accuracy and efficiency of research outcomes.
- Explored the correlation between model accuracy and linguistic features, with a special focus on the challenges posed by Chinese idioms to NLI models, contributing valuable insights into language processing complexities.

Collaborators: Sirui Wu, Tianyu Shao, Yiqi Zhang, Yufan Gai

SELECTED COURSES

Courses in Linguistics at SJTU

- Introduction to English Linguistics
- Research Methods in Linguistics
- Introduction to Syntax
- Phonetics and Phonology
- Introduction to Semantics
- Language Acquisition
- Language Intelligence
- Neurolinguistics

Course in other areas

- Thinking and Methodology in Programming (C++)
- Thinking and Methodology in Programming (Python)
- Calculus
- Linear Algebra
- Probability and Statistics

Yale Summer Session 2023

- Discrete Mathematics
- Introduction to Philosophy

Courses in Progress for Spring 2024

- Data Structures
- Design and Analysis of Algorithms
- Cognitive Psychology and Its Applications
- Machine Learning
- Speech-Language Pathology

ENGLISH PROFICIENCY

TOEFL iBT: 112 (overall score)
Listening: 30 — Reading: 29
Speaking: 26 — Writing: 27
Test date: September 17th, 2023

SKILLS

- **Coding:** Python (pandas,sklearn), C++, LaTeX, Linux
- **Language:** Mandarin(Native), English
- **Other Interests:** Music Making, Lyrics Writing, Cocktail and Coffee Making