

## Skills

---

**Programming:** Python (lib:TensorFlow, PyTorch) R, Matlab, Haskell, Bash, JavaScript, Git, HTML, CSS

**Certification:** Sequence Models | NLP in TensorFlow| TensorFlow for AI | Structuring ML project

**Linguistics:** linguistic research (syntax, semantic, phonetics), grammar development, language data collection

**Languages worked on:** Mandarin, Cantonese, Korean, Japanese, Nepal Bhasa, Somali, English

## Experience

---

**Natural Language Understanding Specialist in Cantonese | *Cerence Inc.*** (currently remote)  
*Jan 2021 - present*

- Annotate structured training data for automotive conversational AI
- Train language models and modify data for quality improvement

**Research Fellow/Teaching Assistant | *Institute of Linguistics at U of M*** *2016 - present*

- Work with community members of low-resource languages, elicit, analyze and store data for research
- Lead various course sections; design lectures, host interactive activities; provide learning feedback

**Computational Linguist Research Intern | *United Language Group*** Minneapolis, MN

- Built parallel corpora for machine translations in Chinese, Japanese, and Korean *Jan - Aug 2017*
- Segmented the data and evaluated the models (rule-and-NN based), using Moses, NLTK, NLPiR

**Language consultant | *Fallon Worldwide*** Minneapolis, MN

- Provided semantic knowledge to the design team, and created a semantic entertainment program *2017*
- Used Twitter API and Python and implemented a program that captures keywords in customers' tweets

**MnDRIVE Scholars | *College of Science and Engineering at U of M*** *May - Aug 2019*

- Designed and taught basic robotic courses to five groups of 30~40 students from Grade 7 to 9
- Taught the courses in programming Ozobots, Inkscape graphic editor, and Intro to virtual reality

## Education

---

**University of Minnesota** **Ph.D. in Linguistics (Computer Science minor)** *August 2021*

Award: Graduate Fellowship

**M.A. in Linguistics** *June 2016*

**Tianjin Foreign Studies University** **B.S. in Educational Technology** *July 2011*

## Projects

---

### Artificial intelligence course project

- Examined whether linguistic factors have effects on the vector space modeling (TensorFlow-word2vec)
- The results show factors have negative effects on vector space modeling due to their syntactic variations

### Computational linguistics course project

- Compared the efficiency of regex taggers and bigram taggers on LCMC corpus, using HMM method
- Regex approach performed better on timesaving, achieved the same accuracy: 85% on unseen data

### MA thesis project

- Examined Entropy Reduction predictions on Mandarin relative clauses, used CTB7 corpus and Tregex tool
- The model predictions reflect the same results from human processing experiments of clausal structures

## Publications

---

First-author publications: Glow-in-Asia XII (2019) | FASAL 7<sup>th</sup> (2017)

Single-author publications: CLS 54<sup>th</sup> (2018) | LSA 92<sup>nd</sup> (2018) | BEAL (2016)

## Volunteers

---

Organizer of Computational Linguistics Group | Officer of Latin Dance Club | Colloquium Student organizer