Borui (Bri) Zhang
zhan3829@umn.edu
https://boruizhang.github.io/

Qualification

- AI techniques: machine learning, machine translation, natural language processing, artificial intelligence, computational linguistics, data structures and algorithms
- Programming skills: Python, JavaScript, Linux, R, Matlab, Haskell, HTML, RegEx, Tregex, git, GitHub; NLP packages and tools: TensorFlow, PyTorch, NLTK, Moses, WordNet, API
- Linguistic research: Syntax and semantics of endangered languages: Newar (Sino-Tibetan), Somali
- Speaking languages: Mandarin, Cantonese, Korean (intermediate level)

NLP/Computational Linguistics Experience

- Computational Linguist Intern at United Language Group, MN *Jan.-Aug. 2017*
  - Built parallel corpora for machine translation projects of the language pairs (zh - en, ja - en, ko - en)
  - Pre-processing and tokenizing data in Linux, using bash, Python, and Vimscript
  - Training MT with different models from rule-based (Moses and NLTK), to neural networks
- Language consultant at Fallon Worldwide, MN *May-Jul. 2016*
  - Created a semantic tool using Twitter API and Python, getting real-time tweets and predicting sentiments
  - Collected and prepared input data as a list of weighted and classified keywords and phrases
  - The program outputs a weighted mixed "flavor" for each customer tweet when any keywords are found
- Computational linguistics project Comparing Unigram Tagger and Bigram Tagger Performance
  - Compared the performance between unigram and bigram taggers, unigram taggers and regular expressions (regex) taggers, and bigram taggers with different backoff methods
  - Used partial LCMC corpus and NLTK, with Hidden Markov methodology
  - A bigram/unigram approach achieved 85% accuracy on unseen data, without any prior knowledge
  - The study found that the bigram/regex method resulting in time savings over the bigram/unigram
- Graduate Teaching Assistant at the University of Minnesota, Twin Cities. *Sep. 2016-current*
  - Preparing linguistics materials (slides and exercises), leading discussions, grading homework and exams
- Web Design Instructor at Asian Media Access Organization, MN *Jan.2019-current*
  - Teaching HTML, CSS, and JavaScript in online/classroom settings
  - Lectures take place in-class meetings, and hands-on practice online, via join.me, CodePen, Google Group

Publications

- [MA thesis] Entropy Reduction Prediction on Mandarin Chinese Relative Clauses, *Buckeye East Asian Linguistics Forum 2016*, The Ohio State University
  - Used Chinese Tree Bank 7.0 and Tregex tool to find all kinds of relative clauses data in the treebank
  - Selected and processed relevant data (relative clauses) with Cornell Conditional Probability Calculator
  - The model predicts the same result as human experiments, that SRCs are easier to process than ORCs
- Sluicing-like construction in Kathmandu Newari *Chicago Linguistics Society* 54th
- Embedding, Covert Movement, and Intervention in Kathmandu Newari *LSA 92nd*
- Processing Embedding Structures in Mandarin Semantics *Workshop of the American Midwest and Prairies*
- Complementation in Newari *Formal Approaches to South Asian Languages Workshop 7th*

Education

- Ph.D. in Linguistics (**Minor in Computer Science**) University of Minnesota *Current*
- M.A. in Linguistics University of Minnesota *June 2016*
- B.S. in Educational Technology Tianjin Foreign Studies University *July 2011*

Members, Awards, & Volunteers

- Member of Women in Computer Science at University of Minnesota *2016-current*
- Leader of Syntax and Semantics Reading Group at University of Minnesota *2018-current*
- Student Organizer of the Linguistics Colloquium at University of Minnesota *2015-2018*
- Granted funding for 27th European Summer School in Logic, Language and Information by European Association for Computer Science Logic (EACSL)