

Zhengxiang Wang

✉ jackwang196531@gmail.com | 🏠 <https://jaaack-wang.eu.org> | 🐙 GitHub | 🔗 LinkedIn

EDUCATION

Stony Brook University

Ph.D. in Linguistics

Stony Brook, NY

Aug 2022 – Present

University of Saskatchewan

M.A. in Applied Linguistics

Saskatoon, Canada

Sep 2019 – May 2021

Hunan University

B.A. in Chinese Language and Literature

Changsha, China

Sep 2015 – Jun 2019

CERTIFICATES

Machine Learning Specialization , DeepLearning.AI	2022
Practical Data Science on the AWS Cloud , DeepLearning.AI	2022
Deep Learning Specialization , DeepLearning.AI	2021
Machine Learning , Stanford Online	2021
Deep Learning Based Natural Language Processing , Baidu PaddlePaddle	2021

PUBLICATIONS & MANUSCRIPTS

(* denotes equal contributions)

- Li, Z*, Makarova, V*. & **Wang, Z***. (2022). Developing Literature Review Writing Skills through an Online Writing Tutorial Series: Corpus-based Evidence. *Frontiers*. Under review.
- **Wang, Z.** (2022). Linguistic Knowledge in Data Augmentation for Natural Language Processing: An Example on Chinese Question Matching. *Proceedings of the Fifth International Conference on Natural Language and Speech Processing* (to appear).
- **Wang, Z.** (2022). Random Text Perturbations Work, but not Always. *Proceedings of the 3rd Workshop on Evaluation and Comparison of NLP Systems*.
- Hao, H., Cui, Y., **Wang, Z.** & Kim, Y. (2022). Thirty-Two Years of IEEE VIS: Authors, Fields of Study and Citations. *Transactions on Visualization and Computer Graphics* (to appear).
- **Wang, Z.** (2021). A Macroscopic Re-examination of Language and Gender: A Corpus-based Case Study in University Instructor Discourses. University of Saskatchewan.
- Peng, L*, & **Wang, Z***. (2017). The Contemporary Reflection on the Sentence-based Theory Postulated by New Grammar of Modern Chinese. *Journal of Hengyang Normal University*. 38(5), 95-101.

PROJECTS

- RNN Seq2seq Models With Attention Are In-distribution Learners** | [GitHub](#) 2022 – Present
 - Designed and conducted comprehensive experiments examining the capabilities of RNN seq2seq models in learning non-regular language transduction tasks for in-distribution & out-distribution test data, given limited training data
- Thirty-two Years of IEEE VIS: Authors, Fields of Study and Citations** | [GitHub](#) | [Manuscript](#) 2022
 - Helped build two text classifiers to predict missing affiliation and country data; analyzed VIS authors over the past 32 years and visualized the collaboration network; contributed to the core idea of visualizing temporal trends
 - Paper accepted by IEEE VIS 2022 and to appear on *Transactions on Visualization and Computer Graphics*
- Deep Learning Based Natural Language Processing using Paddlenlp** | [GitHub](#) 2021 – 2022
 - Finetuned pretrained models, such as BERT, RoBERTa, ERNIE, SKEP, on 11 NLP tasks (e.g., text similarity, sentiment analysis, named entity recognition, relation extraction, Q&A system) to reproduce the SOTA results
- Linguistic Knowledge in Data Augmentation for Natural Language Processing** | [GitHub](#) | [Manuscript](#) 2021
 - Conducted the first (cross-lingual) experiments that demonstrate the limitations of random text perturbations as text augmentation and the minimal role of probabilistic linguistic knowledge in the context of deep learning
 - Designed two text augmentation programs, with or without a N-gram language model, that augment text with 5 token-level text editing operations and can be easily adapted for other languages
- HELPtk: Historical English Language Processing Toolkit** | [GitHub](#) 2021
 - Created a general and open-ended framework that can tokenize, normalize & annotate a normal XML corpus of a few million tokens in few minutes for historical English texts with improved accuracy applying Stanford CoreNLP

- Hand crafted a few hundred normalization and tokenization rules using Regular Expression, informed by the textual distribution of historical English texts discovered by naive Bayes method

A Macroscopic Re-examination of Language and Gender | [GitHub](#) | [Manuscript](#) 2020 – 2021

- Provided the first comparative accounts of male and female university instructor discourses by macroscopically examining the use of 87 syntactic, lexical, and discoursal features in a large compiled corpus
- Developed a general-purpose corpus-linguistic tool to extract and search for linguistic features

TUTORIALS

Text classification & text pair classification explained | [Tutorial1](#) | [Tutorial2](#) 2022

- Building deep learning models for text (pair) classification from scratch using paddle, PyTorch, and TensorFlow

Notes for Stanford CS224N NLP with Deep Learning | [GitHub](#) 2021

- Notes cover the conceptual and mathematical basics of word embedding, neural networks, deep learning models
- Related tutorials of mine: [Word Embedding](#); [Gradient Derivation for ML/DL Loss Functions](#)

EMPLOYMENT

Graduate Teaching Assistant , Stony Brook University	Aug 2022 – Present
Research Mentor , Mitacs	Apr 2021 – Sep 2021
Graduate Research Assistant , University of Saskatchewan	May 2020 – Jun 2021
Research Intern , University of Alberta	Jun 2018 – Sep 2018

SERVICES

Student Organizer , Mathematical Linguistics Reading Group, Stony Brook University	Aug 2022 – Present
Co-founder & President , Hunan University Student Ambassador Association	Apr 2017 – Jul 2019
<ul style="list-style-type: none"> • Helped host and organize dozens of receptions for international guests coming from over ten countries/regions to Hunan University, recruited over 300 volunteers, and raised 15,000 Yuan from partnered organizations 	

SKILLS

Programming: Python, Java, R, Octave, HTML, \LaTeX , C/C++, Unix scripting
Frameworks: TensorFlow, Keras, PyTorch, PaddlePaddle, NumPy, Jax, scikit-learn
Cloud Computing: AWS, Baidu AI Studio, Google Colab

AWARDS & HONORS

Graduate Thesis Award, University of Saskatchewan, 2021
Mitacs Globalink Graduate Fellowship, Mitacs, 2019
Excellent Graduate of the year, Hunan University, 2019
Chinese Government Scholarship, China Scholarship Council, 2018
The First-Class Scholarship, Hunan University, 2017, 2018
National Student Innovation Training Program Grant, Ministry of Education of China, 2017
Triple-A student, Hunan University, 2016
National Scholarship, Ministry of Education of China, 2016