

# 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

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

## PUBLICATIONS & MANUSCRIPTS

(\* denotes equal contributions)

- **Wang, Z.** (2023). Alignment Learning with RNN Seq2seq Models. *Manuscript to appear as a preprint.*
- Li, Z\*, Makarova, V\*. & **Wang, Z\***. (2023). Developing Literature Review Writing Skills through an Online Writing Tutorial Series: Corpus-based Evidence. *Frontiers in Communication-Language Sciences* (to appear).
- **Wang, Z.** (2022). Linguistic Knowledge in Data Augmentation for Natural Language Processing: An Example on Chinese Question Matching. *Proceedings of the 5th International Conference on Natural Language and Speech Processing.*
- **Wang, Z.** (2022). Random Text Perturbations Work, but not Always. *Proceedings of the 3rd Workshop on Evaluation and Comparison of NLP Systems (co-located at ACL 2022).*
- Hao, H., Cui, Y., **Wang, Z.** & Kim, Y. (2022). Thirty-Two Years of IEEE VIS: Authors, Fields of Study and Citations. *IEEE Transactions on Visualization and Computer Graphics.*
- **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

- |  |                |
|--|----------------|
| <b>Alignment Learning with RNN Seq2seq models</b>   <a href="#">GitHub</a>   | 2022 – Present |
| · Designed and conducted comprehensive experiments examining the capabilities of RNN seq2seq models in learning four transduction tasks of varying complexity and that can be described as alignment learning                          |                |
| <b>Thirty-two Years of IEEE VIS: Authors, Fields of Study and Citations</b>   <a href="#">GitHub</a>   <a href="#">Paper</a>   | 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 |                |
| <b>Deep Learning Based Natural Language Processing using Paddlenlp</b>   <a href="#">GitHub</a>  | 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                |                |
| <b>Linguistic Knowledge in Data Augmentation for Natural Language Processing</b>   <a href="#">GitHub</a>   <a href="#">Paper</a>  | 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   |                |
| <b>HELPTk: Historical English Language Processing Toolkit</b>   <a href="#">GitHub</a>   | 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

### rnn-transduction & rnn-seq2seq-transduction | [Tutorial1](#) | [Tutorial2](#) 2023

- Using RNN and RNN seq2seq models (in PyTorch) for modelling string transduction tasks.

### 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](#)

## PRESENTATIONS

- **Wang, Z.** (2023). Understanding how RNN seq2seq models learn alignment. SYNC (STONY BROOK, YALE, NYU, CUNY Linguistics Conference) 23, Yale University.
- **Wang, Z.** (2023). Probing the learning capabilities of RNN seq2seq models (Poster). The 47th Penn Linguistics Conference, University of Pennsylvania.
- **Wang, Z.**, Li, Z., & Makarova, V. (2021) Developing an online academic writing tutorial for non-native English speaking international graduate students in diverse programs of studies. International Teaching Online Symposium, University of Windsor.
- **Wang, Z.**, Tucker, B. V. & Mukai, Y. (2018). How does Chinese learner of English perceive and comprehend phonetic reduction? Summer Poster Symposium at University of Alberta.

## EMPLOYMENT

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

## SERVICES

<b>Student Organizer</b> , Mathematical Linguistics Reading Group, Stony Brook University	Aug 2022 – Present
<b>Co-founder &amp; President</b> , Hunan University Student Ambassador Association	Apr 2017 – Jul 2019
<ul style="list-style-type: none"> <li>• 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</li> </ul>	

## SKILLS

**Programming:** Python, Java, R,  $\text{\LaTeX}$ , Octave, HTML, C/C++, Haskell, 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