

# Zhengxiang Wang

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

## EDUCATION

<b>Stony Brook University</b> <i>Ph.D. in Computational Linguistics</i>	Stony Brook, NY Aug 2022 – Present
<b>University of Saskatchewan</b> <i>M.A. in Applied Linguistics</i>	Saskatoon, Canada Sep 2019 – May 2021
<b>Hunan University</b> <i>B.A. in Chinese Language and Literature</i>	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

## SKILLS

**Programming:** Python, Java, R, L<sup>A</sup>T<sub>E</sub>X, Octave, HTML, C/C++, Haskell, Unix scripting  
**Frameworks:** TensorFlow, Keras, PyTorch, PaddlePaddle, NumPy, Jax, scikit-learn  
**Cloud Computing:** AWS, Baidu AI Studio, Google Colab

## PUBLICATIONS & MANUSCRIPTS

(\* denotes equal contributions)

- **Wang, Z.** (2023). Probabilistic Linguistic Knowledge and Token-level Text Augmentation. *Signals and Communication Technology (Springer)*. Book chapter. In Preparation.
- **Wang, Z.** (2023). Learning Transductions and Alignments with RNN Seq2seq Models. Preprint. *arXiv:2303.06841*.
- 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*.
- **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>RNN Seq2seq Models Learning Transductions and Alignments</b>   <a href="#">GitHub</a>   <a href="#">Manuscript</a>	2022 – 2023
• 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 learning alignments	
<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	

- Linguistic Knowledge in Data Augmentation for Natural Language Processing** | [GitHub](#) | [Paper](#) 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

---

- 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). Exploring RNN Seq2seq Models via Transduction Tasks that Require Alignments. The 10th Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL 2023), George Mason University.
- **Wang, Z.** (2023). Probing the learning capabilities of RNN seq2seq models (Poster). IACS (Institute for Advanced Computational Science) Research Day, Stony Brook University.
- **Wang, Z.** (2023). Probing the learning capabilities of RNN seq2seq models (Poster). The 47th Penn Linguistics Conference, University of Pennsylvania.
- **Wang, Z.** (2023). Understanding how RNN seq2seq models learn alignment. SYNC (STONY BROOK, YALE, NYU, CUNY Linguistics Conference) 23, Yale University.
- **Wang, Z.** (2022). Linguistic Knowledge in Data Augmentation for Natural Language Processing: An Example on Chinese Question Matching. The 5th International Conference on Natural Language and Speech Processing.
- **Wang, Z.** (2022). Random Text Perturbations Work, but not Always. The 3rd Workshop on Evaluation and Comparison of NLP Systems (co-located at ACL 2022).
- Makarova, V., Li, Z. & **Wang, Z.** (2022) What international graduate students seek and find in an online Academic English Writing tutorial course. The Canadian Association of Applied Linguistics (ACLA/CAAL) 2022.
- **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.

## 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