

## EDUCATION

---

### **Stony Brook University**

Ph.D. in Computational Linguistics. Advisors: Owen Rambow, Jeffrey Heinz

Stony Brook, NY

Aug 2022 – Present

### **University of Saskatchewan**

M.A. in Applied Linguistics. Advisor: Zhi Li

Saskatoon, Canada

Sep 2019 – May 2021

### **Hunan University**

B.A. in Chinese Language and Literature

Changsha, China

Sep 2015 – Jun 2019

## INDUSTRY EMPLOYMENT

---

- **Data Scientist Intern**, The Home Depot May 2024 – Aug 2024
  - Developed a clustering-based topic modeling pipeline incorporating LLMs that produces human interpretable and rich-coverage topics while reducing 90+% LLM runtime and API calls and eliminating redundancy in the raw identified topics by 80+%
  - Experimented with several classification methods, such as fine-tuning BERT-like models, fine-tuning SBERT embeddings to build an efficient semantic search network(with Faiss) , fine-tuning generative language models (e.g., full fine-tuning and PEFT such as LoRA) to predict/generate customer pain points
  - Built an LLM-based autonomous validation system with multi-agent debating framework, saving up to 24% manual annotation efforts

## PUBLICATIONS & MANUSCRIPTS

---

### **Book Chapters**

1. Makarova, V., Li, Z. & **Wang, Z.** (2024). Can ChatGPT Grade Non-Native Academic English Writing? In *AI Approaches to Literacy in Higher Education*, page 97-116.
2. **Wang, Z.** (2023). Probabilistic Linguistic Knowledge and Token-level Text Augmentation. In *Practical solutions for Diverse Real-World NLP Applications*, page 1-20. Springer book series *Signals and Communication Technology*.

### **Articles**

1. Li, Z., Makarova, V. & **Wang, Z.** (2023). Developing Literature Review Writing Skills through an Online Writing Tutorial Series: Corpus-based Evidence. *Frontiers in Communication*.
2. 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*.
3. 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*.

Proceedings

1. **Wang, Z.** & Rambow, O. (2024). Clustering Document Parts: Detecting and Characterizing Influence Campaigns From Documents. *Proceedings of the 6th Workshop on Natural Language Processing and Computational Social Science*. (Collocated at NAACL 2024)

2. Li, Z., Makarova, V. & **Wang, Z.** (2023). Assessment of Academic ESL Writing in an Online Tutorial for Graduate Students. *Proceedings of the 22nd European Conference on e-Learning: ECEL 2023*.

3. **Wang, Z.** (2023). Learning Transductions and Alignments with RNN Seq2seq Models. *Proceedings of the 16th International Conference on Grammatical Inference*. In *Proceedings of Machine Learning Research, Volume 217*.

4. **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*.

5. **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)*.

Manuscripts

1. **Wang, Z.**, Kodner, J. & Rambow, O. (2024). Evaluating LLMs with Multiple Problems at once: A New Paradigm for Probing LLM Capabilities. Preprint arXiv 2406.10786.

2. **Wang, Z.** (2021). A Macroscopic Re-examination of Language and Gender: A Corpus-based Case Study in University Instructor Discourses. MA Dissertation, University of Saskatchewan.

3. **Wang, Z.** (2019). Analysis on the quantitateness of reduplication in Modern Mandarin. BA Dissertation (in Chinese), Hunan University.

4. **Wang, Z.** (2017-2019). Grammar as science: Rethinking the construction of Modern Chinese Grammar. Granted project (outcomes are 99 pages/11 papers in Chinese) for the National Student Innovation Training Program.

AWARDS & HONORS

• <b>Junior Researcher Award</b> , Institute for Advanced Computational Science	2024
• <b>Distinguished Travel Award</b> , Graduate Student Organization	2023
• <b>Professional Development Award</b> , Graduate Student Employees Union	2023
• <b>Mitacs Globalink Graduate Fellowship</b> , Mitacs	2019
• <b>Excellent Graduate of the year</b> , Hunan University	2019
• <b>Chinese Government Scholarship</b> , China Scholarship Council	2018
• <b>The First-Class Scholarship</b> , Hunan University	2017, 2018
• <b>National Student Innovation Training Program Grant</b> , Ministry of Education of China	2017
• <b>Triple-A student</b> , Hunan University	2016
• <b>National Scholarship</b> , Ministry of Education of China	2016

RESEARCH EMPLOYMENT

• <b>Research Assistant</b> , under Owen Rambow, SUNY Research Foundation	May 2023 – May 2024 & Aug 2024
DARPA-funded project: Influence Campaigns Awareness and Sensemaking (INCAS).	
• <b>Research Assistant</b> , under Veronika Makarova & Zhi Li, University of Saskatchewan	May 2020 – Jun 2021
Assessing the development of Academic ESL Writing by international graduate students.	
• <b>Research Assistant</b> , under Benjamin V. Tucker, University of Alberta	Summer 2018
Phonological perceptions of Chinese English learners for reduced English sounds.	

## TEACHING EMPLOYMENT

---

- **LIN120: Language and Technology.** Teaching Assistant Fall 2024  
Instructor: Owen Rambow. Stony Brook University.
- **Summer Youth Camp in Computational Linguistics.** Instructor Aug 2023  
Organizer: Jeffrey Heinz, Stony Brook University
- **LIN335: Computational Linguistics.** Teaching Assistant Spring 2023  
Instructor: Jordan Kodner. Stony Brook University.
- **LIN120: Language and Technology.** Teaching Assistant Fall 2022  
Instructor: Owen Rambow. Stony Brook University.

## 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, R, L<sup>A</sup>T<sub>E</sub>X, Octave, HTML, C/C++, Haskell, Bash
- **Frameworks:** PyTorch, TensorFlow, Keras, PaddlePaddle, NumPy, Jax, scikit-learn
- **Cloud Computing:** AWS, Baidu AI Studio, Google Colab
- **Languages:** Mandarin, English, Fuzhou Dialect

## PRESENTATIONS

---

### Talks

1. **Wang, Z.** June 2024. Clustering Document Parts: Detecting and Characterizing Influence Campaigns From Documents. *The 6th Workshop on Natural Language Processing and Computational Social Science*. Collocated at NAACL 2024, Mexico City, Mexico.
2. **Wang, Z.** July 2023. Learning Transductions and Alignments with RNN Seq2seq Models. The 16th International Conference on Grammatical Inference (ICGI), Rabat, Morocco.
3. **Wang, Z.** March 2023. Understanding how RNN seq2seq models learn alignment. SYNC (STONY BROOK, YALE, NYU, CUNY Linguistics Conference) 2023, Yale University.
4. **Wang, Z.** December 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, Online.
5. **Wang, Z.** October 2022. Random Text Perturbations Work, but not Always. The 3rd Workshop on Evaluation and Comparison of NLP Systems (co-located at AACL 2022), Online.

## Posters

1. **Wang, Z.** & Rambow, O. June 2024. Clustering Document Parts: Detecting and Characterizing Influence Campaigns From Documents. The 6th Workshop on Natural Language Processing and Computational Social Science (Collocated with NAACL 2024). Mexico City.
2. **Wang, Z.** April 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.
3. **Wang, Z.** April 2023. Probing the learning capabilities of RNN seq2seq models. IACS (Institute for Advanced Computational Science) Research Day, Stony Brook University.
4. **Wang, Z.** March 2023. Probing the learning capabilities of RNN seq2seq models. The 47th Penn Linguistics Conference, University of Pennsylvania.
5. **Wang, Z.**, Tucker, B. V. & Mukai, Y. August 2018. How does Chinese learner of English perceive and comprehend phonetic reduction? Summer Poster Symposium, University of Alberta.

## SERVICE

---

- Secretary Sep 2024 – Present  
Graduate Student Organization, IACS, Stony Brook University
- Social media manager Aug 2024 – Present  
Linguistics Department, Stony Brook University
- Skillz Hour organizer Sep 2023 – Dec 2023  
Linguistics Department, Stony Brook University
- Student organizer Aug 2022 – May 2023  
Mathematical Linguistics Reading Group, Stony Brook University
- ESL classroom volunteer Dec 2019 – Mar 2020  
The Global Gathering Place, Saskatoon
- Co-founder and president May 2016 – Jun 2019  
Hunan University Student Ambassadors Association, Hunan University
- Student organizer May 2019  
University of Regina's China Study Tour (4 weeks), Hunan University
- Volunteer and interpreter Sep 2018  
UNESCO conference "Role of linguistic diversity" (3 days), Changsha

## AFFILIATION

---

- **Member of Institute for Advanced Computational Science, SBU** Mar 2023 – Present
- **BIAS National Science Foundation Research Traineeship** Jun 2023 – Present
- **Association for Computational Linguistics** Feb 2022 – Present
- **Linguistics Society of America** Apr 2023 – Apr 2024

Last updated on September 12, 2024