Bai Li

Personal Info

Email:	bai@cs.toronto.edu
Website:	www.cs.toronto.edu/~bai/
Citizenship:	Canadian

Education

2019 - 2022 **PhD, Computational Linguistics, University of Toronto**
Thesis: Integrating Linguistic Theory and Neural Language Models
Committee: Frank Rudzicz (Advisor), Yang Xu, Guillaume Thomas

2017 - 2019 **MSc, Computational Linguistics, University of Toronto**
Thesis: Automatic Detection of Dementia in Mandarin Chinese
Advisor: Frank Rudzicz

2012 - 2017 **BMath, Computer Science, University of Waterloo**
French Minor
Faculty Average: 93%

Publications

1. **Li B.**, Zhu Z., Thomas G., Rudzicz F., Xu Y. "Neural reality of argument structure constructions". *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022)*.

2. Zhu Z., Wang J., **Li B.**, Rudzicz F. "On the data requirements of probing". *Findings of the Association for Computational Linguistics (ACL Findings 2022)*.

3. **Li B.**, Zhu Z., Thomas G., Xu Y., Rudzicz F. "How is BERT surprised? Layerwise detection of linguistic anomalies". *Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021)*.

4. **Li B.**, Rudzicz F. "TorontoCL at CMCL 2021 Shared Task: RoBERTa with Multi-Stage Fine-Tuning for Eye-Tracking Prediction". *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics (CMCL at NAACL 2021)*. **Best student paper award**.

5. **Li B.**, Thomas G., Xu Y., Rudzicz F. "Word class flexibility: A deep contextualized approach". *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*.

6. **Li B.**, Xie J.Y., Rudzicz F. "Representation Learning for Discovering Phonemic Tone Contours". *17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology (SIGMORPHON at ACL 2020)*.

7. **Li B.**, Jiang N., Sham J., Shi H., Fazal H. "Real-world Conversational AI for Hotel Bookings". *IEEE Annual Conference on Artificial Intelligence for Industries (AI4I 2019)*. **Awarded best paper of AI4I 2019**.

8. **Li B.**, Hsu Y-T., Rudzicz F. "Detecting dementia in Mandarin Chinese using transfer learning from a parallel corpus". *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)*.

9. Fraser K.C., Linz N., **Li B.**, Fors K.L., Rudzicz F., Konig A., Alexandersson J., Robert P., Kokkinakis D. "Multilingual prediction of Alzheimer's disease through domain adaptation and concept-based language modelling". *2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2019)*.

10. Brzozowski J.A., Kari L., **Li B.**, Szykuła M. "State Complexity of Overlap Assembly". In: Câmpeanu C. (eds) *Implementation and Application of Automata*. CIAA 2018. **Sheng Yu Award for best paper of CIAA 2018**.

Preprints

1. Zhu Z., **Li B.**, Xu Y., Rudzicz F. "What do writing features tell us about AI papers?". *arXiv:2107.06310* (2021).
2. **Li B.**. "Evolution of Part-of-Speech in Classical Chinese". *arXiv:2009.11144* (2020).
3. **Li B.**, Zhang R., Rudzicz F. "Dropout during inference as a model for neurological degeneration in an image captioning network". *arXiv:1808.03747* (2018).

Work Experience

Feb 2022-Jul 2022 | **Founder, LevelText** (Vancouver)
Launched platform for intermediate French learners to find authentic reading material, powered by NLP. Participated in NextAI Incubator (2022 Cohort).

Oct 2021-Apr 2022 | **Senior Machine Learning Scientist, Ada** (Toronto)
Built NLP models for language detection and named entity recognition. Part-time.

Apr 2018-Oct 2020 | **NLP Research Engineer, SnapTravel** (Toronto)
Built NLP system to extract named entities and hotel preferences in a natural language chatbot. Lead author on paper at IEEE AI4I 2019 conference, winning best paper award. Also contributed to machine learning for ad bidding and statistical analysis for experiments. Technologies used: PyTorch, SpaCy, AllenNLP, Elasticsearch, scikit-learn, pandas, R.

Sep-Dec 2016 | **Software Engineer, Yelp** (San Francisco)
Performance team. Built tools for developers to optimize performance, and prototyped a caching proxy service. Technologies: Nginx, Lua, Memcached, Redis.

Jan-Apr 2016 | **Software Engineer, Yext** (New York)
Core Platform team. Technologies: Java, Play Framework, Hibernate, Protocol Buffers, Groovy / Soy templates, Docker.

May-Aug 2015 | **Software Engineer, Minted** (San Francisco)
Developed system for designers and artists to launch digital stationery products to an online marketplace. Technologies: AWS, SQLAlchemy, Bottle, Backbone.js.

Sep-Dec 2014 | **Software Engineer, A Thinking Ape** (Vancouver)
Full stack engineer for top-selling mobile games Kingdoms at War (KaW) and Party In My Dorm (PIMD). Technologies: Django Framework (Python), Android, MySQL, Redis, Websocket.

Jan-Apr 2014 | **Software Developer, TutorJam Canada** (Kitchener)
Worked on YuJa, an online video collaboration platform. Technologies: Java EE, Wildfly, jQuery, SQL, Bootstrap, AWS, Twilio.

Teaching Experience

1. CSC 2541 (Toronto) – Machine Learning for Healthcare – TA – Winter 2019
2. CSC 485/2501 (Toronto) – Computational Linguistics – TA – Fall 2018
3. CSC 401/2511 (Toronto) – Natural Language Processing – TA – Winter 2018
4. CSC 236 (Toronto) – Introduction to Theory of Computation – TA – Fall 2017
5. MATH 145 (Waterloo) – Algebra (Advanced Level) – TA – Fall 2014

Side Projects

- **WATisRain**
  Android and iOS app to help you navigate the buildings of the University of Waterloo without going outside. Installed by thousands of Waterloo students, with 4.6 of 5 rating on Google Play. Github project at https://github.com/luckytoilet/WATisRain .
- **Lucky's Notes Blog**
  Explains complicated topics in math and computer science in plain English. Available at https://luckytoilet.wordpress.com.

Interests and Hobbies

- Language learner: English (native), Mandarin Chinese (native), French (B2), Spanish (B1), Japanese (B1), Teochew (B1)
- Writes a Math and Computer Science blog (http://luckytoilet.wordpress.com)
- Loves reading and learning anything and everything
- Plays piano and guitar; makes own arrangements for anime and pop songs

Scholarships and Awards

2018-2019	Queen Elizabeth II Graduate Scholarship in Science & Technology (QEII-GSST)
2018	Sheng Yu Award for Best Paper at CIAA 2018
2018	Vector Institute Research Award
2012-2017	University of Waterloo Faculty of Mathematics Scholarship
2016	University of Waterloo President's Research Award
2014-2015	University of Waterloo President's International Experience Award
2012-2013	University of Waterloo President's Scholarship of Distinction