

Sang Yun Kwon

PH.D. STUDENT · SCHOOL OF INFORMATION · UNIVERSITY OF BRITISH COLUMBIA (UBC)

2075, West Mall, Vancouver, BC, Canada V6T 1Z2

☎ (+1) 604-773-6305 | ✉ skwon01@mail.ubc.ca | 🏠 skwon01-ubc.github.io/website/ | 🔗 [LinkedIn](#) | 🎓 [Google Scholar](#)

Research Interests

My research interests lie in natural language processing and machine learning, with a focus on cultural representation and inclusivity in large language models. I am particularly interested in developing technologies that serve underrepresented language communities and bridge cross-cultural gaps in NLP systems.

Education

University of British Columbia (UBC)

Vancouver, BC, Canada

PH.D. IN INFORMATION

Sep. 2024 - Nov. 2028 (Expected)

- Advisor: [Prof. Muhammad Abdul-Mageed](#)
- GPA: 4.00/4.00

University of British Columbia (UBC)

Vancouver, BC, Canada

M.A. LIBRARY AND INFORMATION STUDIES

Sep. 2021 - Aug. 2024

- Advisor: [Prof. Muhammad Abdul-Mageed](#)
- GPA: 4.00/4.00

Yonsei University (YU)

Seoul, South Korea

B.A. INTERNATIONAL COMMERCE

Match. 2015 - Aug. 2020

- GPA: 3.30/4.00

Publications

CONFERENCE PUBLICATIONS

- [C.4] AbdelRahim Elmadany, [Sang Yun Kwon](#)^{*}, FaHawau Olamide Toyin, Alcides Alcoba Inciarte, Hanan Aldarmaki, and Muhammad Abdul-Mageed. “**Voice of a Continent: Mapping Africa’s Speech Technology Frontier**”. *The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Nov. 2025.
- [C.3] Samar Magdy^{*}, [Sang Yun Kwon](#)^{*}, Fakhraddin Alwajih, Safaa Taher Abdelfadil, Shady Shehata, and Muhammad Abdul-Mageed. “**JAWA-HER: A Multidialectal Dataset of Arabic Proverbs for LLM Benchmarking**”. *Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL)*. Nov. 2025.
- [C.2] Samar Magdy, [Sang Yun Kwon](#), Fakhraddin Alwajih, Reem Abdel-Salam, and Muhammad Abdul-Mageed. “**Gazelle: An Instruction Dataset for Arabic Writing Assistance**”. *The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Aug. 2024. [\[Arxiv\]](#)
- [C.1] [Sang Yun Kwon](#), Gagan Bhatia, El Moatez Billah Nagoudi, and Muhammad Abdul-Mageed. “**Beyond English: Evaluating LLMs for Arabic Grammatical Error Correction**”. *Arabic Natural Language Processing Conference (ArabicNLP)*. Aug. 2023. [\[Arxiv\]](#)

WORKSHOP PUBLICATIONS AND POSTERS

- [W.1] [Sang Yun Kwon](#), Gagan Bhatia, El Moatez Billah Nagoudi, Alcides Alcoba Inciarte, and Muhammad Abdul-Mageed. “**SIDLR: Slot and Intent Detection Models for Low-Resource Language Varieties**”. *Tenth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial’23)*. May. 2023. [\[Arxiv\]](#)

Honors and Awards

2021-2023 Graduate Scholarship of School of Information

Canada

2023 Masters Travel Award of School of Information, University of British Columbia (UBC)

Canada

2018-2019 Yonsei University Academic Honors (Top 25% in Class)

South Korea

Experience

Master of Data Science in Computational Linguistics

TEACHING ASSISTANT

- DSCI-511 Python Programming

Vancouver, BC, Canada

Sept. 2025 - Present

UBC Deep Learning and Natural Language Processing Group, University of British Columbia (UBC)

GRADUATE RESEARCH ASSISTANT

- Advisor: Prof. Muhammad Abdul-Mageed
- Research in NLP, Grammatical Error Correction, Text generation, and Language modeling.
- Creative text generation and collaborative writing.
- Developed a state-of-the-art (SoTA) Grammatical Error Correction model specifically tailored for the Arabic language.
- Engaged in language modeling and dataset creation focused on Slot and Intent Detection for Low-resource Languages.
- Publishing papers in top-tier NLP venues and delivering presentations.

Vancouver, BC, Canada

June. 2022 - Present

Talks

Apr. 2025 **Guest Lecture: LLM Agents**, UBC, Master of Data Science in Computational Linguistics

Vancouver

Dec. 2023 **Beyond English: Evaluating LLMs for Arabic Grammatical Error Correction**, at ArabicNLP’23

Singapore

Service

Aug. 2024 **Reviewer**, at ArabicNLP’25

Suzhou, China

Aug. 2024 **Reviewer**, at ArabicNLP’24

Bangkok, Thailand

Dec. 2023 **Reviewer**, at ArabicNLP’23

Singapore

Skills

Programming Languages | Tools R, Python, Pytorch, DSPy, LangChain, CrewAI

Relevant Coursework

- Applied Machine Learning, UBC
 - Machine Learning and Data Mining, UBC
- Designing Interactive Computational Technology for People, UBC
 - Topics in HCI, UBC
- Database Design, UBC
 - Python Programming, UBC