

Haiyue Song

Updated June 21, 2024

Homepage: <https://shyyhs.github.io>
Email: haiyue.song@nict.go.jp
Position: Technical Researcher
National Institute of Information and Communications Technology (NICT)
Address: 3-5 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0289, Japan

Research Interests Machine translation (MT), including low-resource MT, domain adaptation in MT, multi-modal MT, linguistically motivated MT, and LLMs for MT.
Subword, including subword segmentation, encoding, and decoding.

Education

Kyoto University Kyoto, Japan
Ph.D. in Intelligence Science and Technology Oct. 2020 – Mar. 2024
Thesis: *Studies on Subword-based Low-Resource Neural Machine Translation: Segmentation, Encoding, and Decoding*
Advisors: Sadao Kurohashi, Chenhui Chu

Kyoto University Kyoto, Japan
Master of Intelligence Science and Technology Oct. 2018 – Sept. 2020
Advisors: Sadao Kurohashi, Chenhui Chu

Shanghai Jiao Tong University Shanghai, China
Bachelor of Computer Science and Technology Sept. 2014 – July 2018
Minor in Japanese Mar. 2016 – July 2018
Advisor: Li Jiang

Honors and Awards **Research Fellowship for Young Scientists (DC1)** Apr. 2021 – June 2023
Japan Society for the Promotion of Science

Research Experience **Technical Researcher, NICT**
Advisors: Masao Utiyama, Hideki Tanaka, Raj Dabre July 2023 – Present
Research on machine translation including MT for data with document structure, low-resource MT, LLMs for MT, and speech translation

Research Internship, NICT
Advisors: Masao Utiyama, Hideki Tanaka, Raj Dabre Oct. 2019 – June 2023
Research on low-resource machine translation and subword segmentation

Research Assistant, Kyoto University
Advisors: Sadao Kurohashi Nov. 2020 – Mar. 2021
Project related to machine translation

Internship, Machine Learning Team, LINE Tokyo, Japan

Selected Publications

Bilingual Corpus Mining and Multistage Fine-Tuning for Improving Machine Translation of Lecture Transcripts

Haiyue Song, Raj Dabre, Chenhui Chu, Atsushi Fujita, and Sadao Kurohashi.
Journal of Information Processing, Accepted to Journal of Information Processing

DiverSeg: Leveraging Diverse Segmentations with Cross-granularity Alignment for Neural Machine Translation

Haiyue Song, Zhuoyuan Mao, Raj Dabre, Chenhui Chu, and Sadao Kurohashi.
Journal of Natural Language Processing, 2024 Volume 31 Issue 1 Pages 155–188

SelfSeg: A Self-supervised Sub-word Segmentation Method for Neural Machine Translation

Haiyue Song, Raj Dabre, Chenhui Chu, Sadao Kurohashi, and Eiichiro Sumita.
ACM Trans. Asian Low-Resour. Lang. Inf. Process. (2023.7)

SubMerge: Merging Equivalent Subword Tokenizations for Subword Regularized Models in Neural Machine Translation

Haiyue Song, Francois Meyer, Raj Dabre, Hideki Tanaka, Chenhui Chu, and Sadao Kurohashi.

Accepted to The 25th Annual Conference of the European Association for Machine Translation (EAMT 2024)

BERTSeg: BERT Based Unsupervised Subword Segmentation for Neural Machine Translation

Haiyue Song, Raj Dabre, Zhuoyuan Mao, Chenhui Chu, and Sadao Kurohashi.
Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (ACL 2022)

Patents

One patent application about subword segmentation in progress 2022

Talks and Tutorials

Tutorial: Linguistically Motivated Neural Machine Translation 2024
The 25th Annual Conference of the European Association for Machine Translation (EAMT 2024)

Service

Reviewer of TASLP2024, ARR2024, TALLIP2024, TALLIP2023, ARR2023, APSIPA ASC2023, EMNLP2023, ACL2023, EMNLP2022, EMNLP2021, EMNLP2020, IJCNLP2020, and WAT2020

Mentor for intern students at NICT

Mentor for ACL2020-SRW