WAT 2025

The 12th Workshop on Asian Translation

Proceedings of the Workshop

December 24, 2025

Order copies of this and other ACL proceedings from:

Preface

Many Asian countries are rapidly growing these days and the importance of communicating and exchanging the information with these countries has intensified. To satisfy the demand for communication among these countries, machine translation technology is essential.

Machine translation technology has rapidly evolved recently and it is seeing practical use especially between European languages. However, the translation quality of Asian languages is not that high compared to that of European languages, and machine translation technology for these languages has not reached a stage of proliferation yet. This is not only due to the lack of the language resources for Asian languages but also due to the lack of techniques to correctly transfer the meaning of sentences from/to Asian languages. Consequently, a place for gathering and sharing the resources and knowledge about Asian language translation is necessary to enhance machine translation research for Asian languages.

The Conference on Machine Translation (WMT), the world's largest machine translation conference, mainly targets on European language. The International Workshop on Spoken Language Translation (IWSLT) has spoken language translation tasks for some Asian languages using TED talk data, but there is no task for written language. The Workshop on Asian Translation (WAT) is an open machine translation evaluation campaign focusing on Asian languages. WAT gathers and shares the resources and knowledge of Asian language translation to understand the problems to be solved for the practical use of machine translation technologies among all Asian countries. WAT is unique in that it is an open innovation platform": the test data is fixed and open, so participants can repeat evaluations on the same data and confirm changes in translation accuracy over time. WAT has no deadline for the automatic translation quality evaluation (continuous evaluation), so participants can submit translation results at any time.

Following the success of the previous WAT workshops (WAT2014 – WAT2024), WAT2025 will bring together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 12th WAT, we have 1 Patent Translation Task, 2 Document Translation Tasks and 5 Multimodal Translation Tasks. We have 6 teams who submitted their results.

In addition to the shared tasks, WAT2025 also features research papers on topics related to machine translation, especially for Asian languages. We received 7 research papers submitted including ARR commitment, and the program committee accepted 4 research papers.

We would like to thank all the authors who submitted papers. We express our deepest gratitude to the committee members for their timely reviews. We also thank the AACL-IJCNLP 2025 organizers for their help with administrative matters.

WAT 2025 Organizers

Organizing Committee

Organizers

Toshiaki Nakazawa, The University of Tokyo, Japan
Isao Goto, Ehime University, Japan
Hidaya Mino, Japan Broadcasting Corporation (NHK), Japan
Naoto Shirai, Japan Broadcasting Corporation (NHK), Japan
Chenhui Chu, Kyoto University, Japan
Haiyue Song, National Institute of Information and Communications Technology (NICT), Japan
Raj Dabre, Google Research Australia, Australia
Shohei Higashiyama, National Institute of Information and Communications Technology (NICT), Japan
Anoop Kunchookuttan, Microsoft AI and Research, India
Shantipriya Parida, AMD Silo AI, Finland
Ondřej Bojar, Charles University, Prague, Czech Republic
Katsuhito Sudoh, Nara Women's University, Japan
Masaaki Nagata, NTT Corporation, Japan
Sadao Kurohashi, National Institute of Informatics, Japan

Technical Collaborators

Luis Fernando D'Haro, Universidad Politécnica de Madrid, Spain
Rafael E. Banchs, Nanyang Technological University, Singapore
Haizhou Li, National University of Singapore, Singapore
Chen Zhang, National University of Singapore, Singapore

Program Committee

Program Committee Members

Raj Dabre, Google Research Australia, Australia
Shohei Higashiyama, NICT
Chao-Hong Liu, Potamu Research Limited
Hideya Mino, NHK
Takashi Ninomiya, Ehime University
Shantipriya Parida, AMD Silo AI
Katsuhito Sudoh, Nara Women's University
Masao Utiyama, NICT

<center>

Keynote Talk

Optimizing Large Language Models for Low-resource Quality Estimation

</center>

<center>

Diptesh Kanojia
University of Surrey
2025-12-24 11:40:00 –

</center>

Abstract: Large Language Models (LLMs) are positioned as generalist models often claiming superlative performance on many Natural Language Processing (NLP) tasks. However, they tend to fail at Quality Estimation (QE) of Machine Translation (MT), particularly for low-resource languages. The talk investigates root causes of this disparity, such as tokenization inconsistencies arising from morphological richness in natural languages. To bridge this gap, the talk introduces strategies to embed annotation guidelines-based reasoning constraints directly in-context. Furthermore, our investigation on optimal cross-lingual alignment shows that intermediate Transformer layers help produce performant adapters. By attaching Low-Rank Adapter (LoRA) based regression heads to intermediate layers, we bypass the generation-specific biases of the final layer, efficiently outperforming standard instruction fine-tuning and SoTA encoders like COMETKiwi. Finally, via results from the WMT Unified Shared subtask on QE-informed Correction, we demonstrate that these precise estimations can guide LLMs to produce reliable corrections. We discuss how these signals help address the diminishing returnschallenge, enabling models to improve fluent outputs without diverging from human references.

Bio: Researcher working on problems within areas of Natural Language Processing (NLP) and Machine Learning (ML) at the Institute for People-Centred AI (PAI) and School of Computer Science and Electronic Engineering. As a research lead, I manage the NLP subgroup within the Nature Inspired Computing and Engineering group (NICE) @ Computer Science Research Centre. I also lead teaching on the NLP module offered to both undergraduate and postgraduate students.

My research focuses on developing scalable and safe human-machine interaction using foundation models. Guided by the principles of Responsible and Inclusive AI, my work emphasises cross-lingual and multimodal representation learning to address challenges like online toxicity, misinformation, and digital accessibility for low-resource languages. Our research outcomes- code, data, and models, are publicly available on the SurreyNLP GitHub and HuggingFace.

My prior roles include a Postdoctoral Fellowship at the Centre for Translation Studies, a joint PhD from IIT Bombay and Monash University, and Research Engineer at the CFILT Lab.

Table of Contents

Program

CycleDistill: Bootstrapping Machine Translation using LLMs with Cyclical Distillation
Deepon Halder, Thanmay Jayakumar and Raj Dabre

14:50 - 15:30   *Shared Task - Japanese-English Article-level News Translation*

Findings of the WAT 2025 Shared Task on Japanese-English Article-level News Translation
Naoto Shirai, Kazutaka Kinugawa, Hitoshi Ito, Hideya Mino and Yoshihiko Kawai

NHK Submission to WAT 2025: Leveraging Preference Optimization for Article-level Japanese–English News Translation
Hideya Mino, Rei Endo and Yoshihiko Kawai

15:30 - 16:00   *Break*

16:00 - 17:20   *Shared Task - English-to-Indic Multimodal Translation*

Findings of WAT2025 English-to-Indic Multimodal Translation Task
Shantipriya Parida and Ondřej Bojar

OdiaGenAI participation at WAT 2025
Debasish Dhal, Sambit Sekhar, Revathy V R, Shantipriya Parida and Akash Kumar Dhaka

Does Vision Still Help? Multimodal Translation with CLIP-Based Image Selection
Deepak Kumar, Baban Gain, Kshetrimayum Boynao Singh and Asif Ekbal

A Picture is Worth a Thousand (Correct) Captions: A Vision-Guided Judge-Corrector System for Multimodal Machine Translation
Siddharth Betala, Kushan Raj, Vipul Betala and Rohan Saswade

17:20 - 17:25   *Closing*