Simultaneous Speech Translation in Distant Language Pairs incorporating Human Interpreters' Translation Strategies

Name: Mana Makinae

Laboratory's name: Natural Language Processing Laboratory

Supervisor's name: Taro Watanabe

Abstract (should be within 1st page)

Simultaneous speech translation (SimulST) aims to translate incoming source speech incrementally while balancing translation quality and latency. Although previous studies have investigated this trade-off, existing approaches often rely on offline translation data and evaluation methods that may not adequately reflect SimulST constraints. This dissertation addresses this gap by leveraging large language models (LLMs) and human simultaneous interpretation strategies, particularly monotonicity and omission. First, we construct Simul-MuST-C, a dataset using LLMs to align target translations with source speech through minimal interpretable chunks. Experiments show that this approach improves the quality-latency trade-off, particularly for language pairs with substantial word-order differences. Second, we investigate omission by constructing omission-aware training data and proposing Target-Duration Latency (TDL), a metric accounting for target output length. Results show that omission can shorten translations while preserving essential information, although conventional evaluation metrics may overlook its benefits. Third, we introduce simul-tst-COMMON, a test set designed to maintain source-target word and phrase order consistency, with translation quality verified by professional simultaneous interpreters. Experiments reveal that offline-style references may underestimate SimulST performance, highlighting the importance of simultaneous-oriented evaluation data. Finally, we propose Simul-COMET-QE, a reference-free quality metric incorporating monotonic translation characteristics by adapting existing training data and quality scores. The metric improves the preference for interpreter-verified simultaneous translations over offline-style translations, particularly for English--Japanese. Together, these findings highlight the importance of adapting translation style and evaluation to SimulST constraints, particularly for language pairs with substantial word-order differences.