Investigating the Influence of Multi-modal input on Human Evaluation of Machine Translation Quality

Sami Ul Haq
Dublin City University

Sheila Castilho
Dublin City University

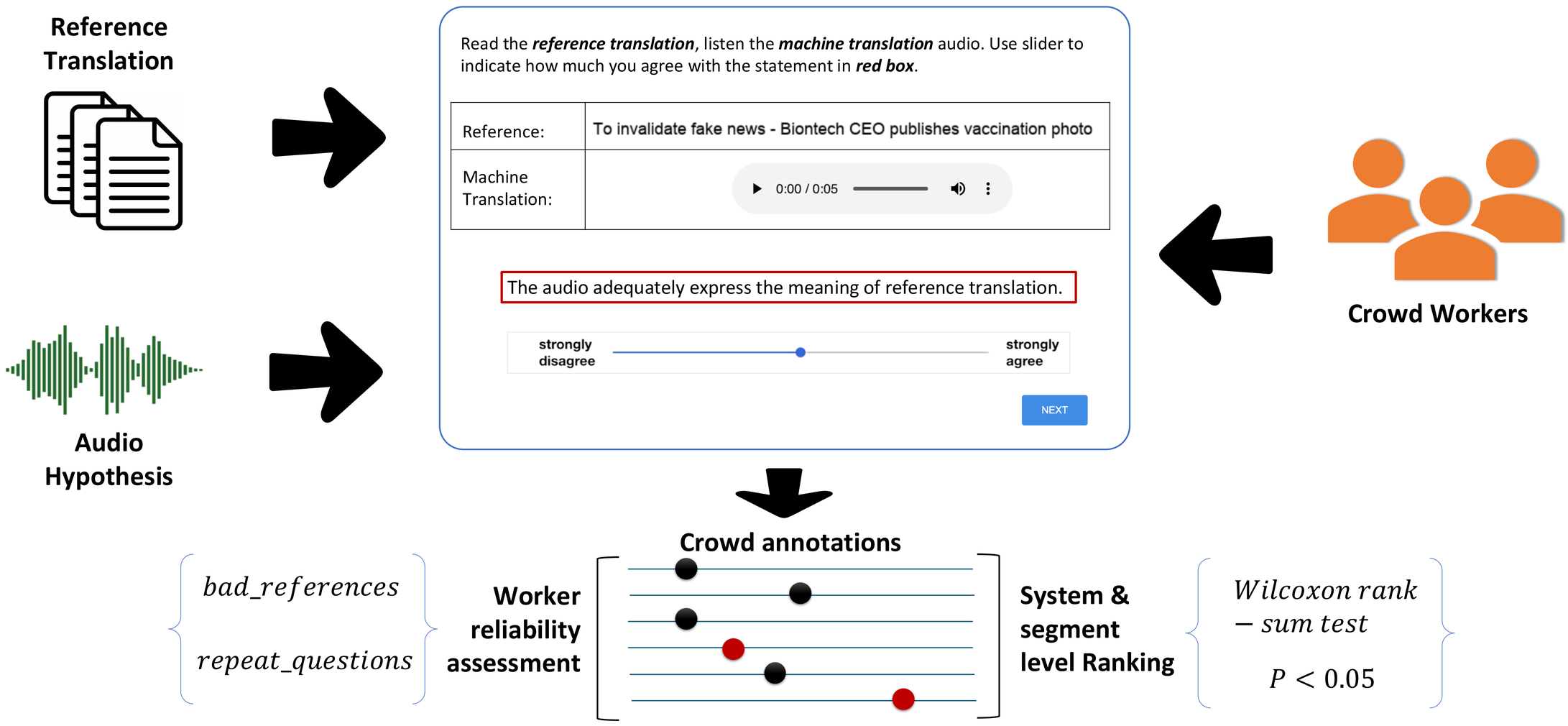
Yvette Graham
Trinity College Dublin

Why Human Evaluation is Important?

- Human feedback stands as the gold standard for assessing machine translation (MT) quality, aligning closely with users' actual needs and expectations.
- Ensuring the accuracy and adequacy of the translated text is crucial in preventing misunderstandings, errors, and potential legal or financial repercussions.
- While automatic evaluation metrics offer a quick substitute for measurements, their correlation with human judgments is weak.
- Expert-based assessment of translation quality is both expensive and challenging to implement on large-scale.

Multi-modal Input

- Multi-modal MT can directly translate to/from multiple modalities (such as Image, Text, Audio).
- Multi-modal evaluation metrics are essential for comprehensive assessment of MT across different modalities.
- Introducing additional technologies such as Text-to-Speech (TTS) to assessment environments may help identify subtle MT errors as compared to text-only conditions.
- Can automatic speech synthesis be leveraged to enhance MT quality assessment?
- Which of two conditions (text only, or target sound) is more conducive to better assessment of MT quality?



Crowd-sourcing Human Judgements

- Access to large pool of native speakers of different languages from around the world.
- Reference based Direct Assessment (DA) does not require bilingual experts.
- Low-cost assessments provide a viable means for individual MT developers and researchers to assess system improvements over a baseline.

Forthcoming

- Crowd annotation collection and analysis to investigate the of impact of sound conditions in comparison to text-only scenarios.
- Enabling natural-sounding artificial voices and flexibility to select between silent or sound conditions.
- Exploring the integration of Visual modality (Images, videos) to provide extra contextual information about translation.

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at DCU [13/RC/2106_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.