ELITR: European Live Translator

Ondřej Bojar[1]   **Dominik Macháček**[1]   **Sangeet Sagar**[1]   **Otakar Smrž**[1]
Jonáš Kratochvíl[1]   **Peter Polák**[1]   **Ebrahim Ansari**[1]
Dario Franceschini[2]   **Chiara Canton**[2]   **Ivan Simonini**[2]
Thai-Son Nguyen[3]   **Felix Schneider**[3]   **Sebastian Stüker**[3]   **Alex Waibel**[3]
Barry Haddow[4]   **Rico Sennrich**[4]   **Philip Williams**[4]
[1]Charles University, [2]PerVoice, [3]Karlsruhe Institute of Technology, [4]University of Edinburgh
Coordinator email: `bojar@ufal.mff.cuni.cz`

Abstract

ELITR (European Live Translator) project aims to create a speech translation system for simultaneous subtitling of conferences and online meetings targetting up to 43 languages. The technology is tested by the Supreme Audit Office of the Czech Republic and by alfaview®, a German online conferencing system. Other project goals are to advance document-level and multilingual machine translation, automatic speech recognition, and meeting summarization.

1 Description

ELITR (European Live Translator, `elitr.eu`) is a three-year EU H2020 Research and Innovation Programme running from 2019 to 2021. The consortium consists of Charles University, University of Edinburgh, Karlsruhe Institute of Technology (research partners), PerVoice (integrator) and alfatraining (user partner).

2 Objectives

ELITR objectives are research and innovations in the field of spoken language and text translation and automatic summarization of meetings.

2.1 Simultaneous Subtitling

In ELITR, we aim to develop a system for simultaneous subtitling of conferences and online meetings. Our affiliated user partner is the Supreme Audit Office of the Czech Republic. It is hosting a congress of EUROSAI (European Organization of Supreme Audit Institutions). The congress participants are natives of 43 languages, and many of them have difficulties in understanding any of the six congress official languages, into which it is interpreted by humans, or to understand some non-native accents. For this and other similar cases, we develop a simultaneous speech translation system from 7 spoken languages (English, German, Russian, Italian, French, Spanish, and experimentally Czech) subtitling into 43 languages, including those for which a human interpreter would not be available for capacity reasons. The 43 languages are 24 EU official languages and 19 others, spoken between Morocco and Kazakhstan.

With our other user partner, alfatraining, we connect our system with an online meeting platform, alfaview®.

2.2 Other Research Topics

The most visible application goal of live subtitling is supported by our advancements in the related areas. We research into document-level machine translation to enable conference participants to translate documents between all the 43 languages in high-quality, taking inter-sentential phenomena into account (Voita et al., 2019a; Voita et al., 2019b; Vojtěchová et al., 2019; Popel et al., 2019; Rysová et al., 2019).

We research into multilingual machine translation to reduce the cost of targeting many languages at once, and to leverage multiple language variants of the source for higher quality (Zhang et al., 2019; Zhang and Sennrich, 2019).

To face challenges of simultaneous translation, such as robustness to noise, out-of-vocabulary words, domain adaptation, and non-standard accents (Macháček et al., 2019), latency and quality trade-off, we aim to improve automatic speech

recognition. We also explore cascaded and fully end-to-end neural spoken language translation (Pham et al., 2019; Nguyen et al., 2019; Nguyen et al., 2020) and co-organize shared tasks at WMT and IWSLT.

2.3 Automatic Minuting

The last objective of our project is an automatic system for structured summaries of meetings. It is a challenging and high-risk goal, but potentially very profitable. We aim to lay the necessary foundations for research in this area by collecting and releasing relevant datasets (Nedoluzhko and Bojar, 2019; Çano and Bojar, 2019a; Çano and Bojar, 2019b) and plan to run shared tasks.

3 ELITR SLT System

ELITR's integration of components of spoken language translation builds on a proprietary software solution by the project integrator PerVoice. The central point is a server called the "mediator". "Workers" for SLT subtasks, such as automatic speech recognition, machine translation, and intermediate punctuating component for specific languages, potentially in event-specific or experimental versions, are provided by research labs in the consortium, ran on their hardware and connected to the mediator. A client requests a specific task, for example, German audio into Czech translation, and the mediator connects a cascade of workers to deliver the requested output. The last worker finally publishes subtitles on a webpage. Meeting participants follow subtitles and slides on personal devices.

The system provides simultaneous low latency translation. We follow the re-translation approach of Niehues et. al (2018). The translation is first displayed around 1 second after the speaker, and then it is occasionally corrected and finalized after approximately 7 seconds.

Acknowledgement

References

Çano, Erion and Ondřej Bojar. 2019a. Efficiency Metrics for Data-Driven Models: A Text Summarization Case Study. *arXiv e-prints*, page arXiv:1909.06618.

Çano, Erion and Ondřej Bojar. 2019b. Keyphrase generation: A text summarization struggle. In *NAACL/HLT*, June.

Macháček, Dominik, Jonáš Kratochvíl, Tereza Vojtěchová, and Ondřej Bojar. 2019. A Speech Test Set of Practice Business Presentations with Additional Relevant Texts. In *SLSP*.

Nedoluzhko, Anna and Ondřej Bojar. 2019. Towards Automatic Minuting of Meetings. In *ITAT: Slovenskočeský NLP workshop (SloNLP 2019)*.

Nguyen, Thai-Son, Sebastian Stueker, Jan Niehues, and Alex Waibel. 2019. Improving sequence-to-sequence speech recognition training with on-the-fly data augmentation. *arXiv preprint arXiv:1910.13296*.

Nguyen, Thai-Son, Sebastian Stüker, and Alex Waibel. 2020. Toward Cross-Domain Speech Recognition with End-to-End Models. *arXiv preprint arXiv:2003.04194*.

Niehues, Jan, Ngoc-Quan Pham, Thanh-Le Ha, Matthias Sperber, and Alex Waibel. 2018. Low-latency neural speech translation. *arXiv preprint arXiv:1808.00491*.

Pham, Ngoc-Quan, Thai-Son Nguyen, Jan Niehues, Markus Müller, and Alex Waibel. 2019. Very Deep Self-Attention Networks for End-to-End Speech Recognition. *Proc. Interspeech, Graz, Austria*.

Popel, Martin, Dominik Macháček, Michal Auersperger, Ondřej Bojar, and Pavel Pecina. 2019. English-Czech Systems in WMT19: Document-Level Transformer. In *WMT Shared Task Papers*.

Rysová, Kateřina, Magdaléna Rysová, Tomáš Musil, Lucie Poláková, and Ondřej Bojar. 2019. A Test Suite and Manual Evaluation of Document-Level NMT at WMT19. In *WMT Shared Task Papers*.

Voita, Elena, Rico Sennrich, and Ivan Titov. 2019a. Context-Aware Monolingual Repair for Neural Machine Translation. In *EMNLP/IJCNLP*.

Voita, Elena, Rico Sennrich, and Ivan Titov. 2019b. When a Good Translation is Wrong in Context: Context-Aware Machine Translation Improves on Deixis, Ellipsis, and Lexical Cohesion. In *ACL*.

Vojtěchová, Tereza, Michal Novák, Miloš Klouček, and Ondřej Bojar. 2019. SAO WMT19 Test Suite: Machine Translation of Audit Reports. In *WMT Shared Task Papers*.

Zhang, Biao and Rico Sennrich. 2019. Root Mean Square Layer Normalization. In *NIPS*, Vancouver, Canada.

Zhang, Biao, Ivan Titov, and Rico Sennrich. 2019. Improving Deep Transformer with Depth-Scaled Initialization and Merged Attention. In *EMNLP/IJCNLP*.