Simultaneous Translation

Liang Huang [1,2]   **Colin Cherry** [3]   **Mingbo Ma** [2]   **Naveen Arizhabagan** [3]   **Zhongjun He** [4]

[1]Oregon State University   [2]Baidu Research USA   [3]Google, Inc.   [4]Baidu, Inc.

1 Brief Description

Simultaneous translation, which performs translation concurrently with the source speech, is widely useful in many scenarios such as international conferences, negotiations, press releases, legal proceedings, and medicine. This problem has long been considered one of the hardest problems in AI and one of its holy grails. Recently, with rapid improvements in machine translation, speech recognition, and speech synthesis, there has been exciting progress towards simultaneous translation. This tutorial will focus on the design and evaluation of policies for simultaneous translation, to leave attendees with a deep technical understanding of the history, the recent advances, and remaining challenges in this field.

2 Type of the Tutorial

This is a **cutting-edge** proposal, and the **first tutorial** on this topic (simultaneous translation) in the history of ACL, EMNLP, NAACL, EACL, COLING, and AACL.

3 Outline

- Background: Simultaneous Interpretation (15 min.)

- Overview of Challenges and Existing Approaches to Simultaneous Translation (25 min.)
    - tradeoff between quality and latency
    - drastic word orders difference
    - robustness, such as error propagation

- Prefix-to-Prefix Framework and Fixed-Latency Policies (15 min.)

- Latency Metrics (20 min.)
    - Average Proportion (AP)
    - Consecutive Wait (CW)
    - Average Lagging (AL)
    - Differentiable Average Lagging (DAL)
    - Ear-to-Voice Span (EVS)

- Dynamic Policies, Part I (15 min.)
    - Adaptive policy with manually designed criteria
    - Reinforcement Learning-based methods
    - Supervised policy-learning framework

- (Coffee Break)

- Dynamic Policies, Part II: Recent Advances (30 min.)
    - Monotonic Infinite Lookback attention
    - Context-Aware translation

- Dataset for Training and Evaluating Simultaneous Translation (20 min.)
    - Rewriting (paraphrasing) references of parallel text
    - Simultaneous Translation datasets:
        * UN corpus
        * EPIC corpus
        * NAIST dataset
        * BSTC dataset

- Towards Simultaneous Speech-to-Speech Translation (20 min.)

- Practical System and Products (20 min.)
    - Practical Issues (segmentation, punctuation, error tolerance)
    - speech-to-text and speech-to-speech systems
    - computer aided interpretation (CAI)

4 Breadth

We envision a tutorial that emphasizes interdisciplinary breadth at the beginning and end (roughly one half of the tutorial in total). The beginning section on Human Interpretation will allow us to discuss the strategies and behaviours that enable humans to perform this challenging task, touching on observations from Translation Studies. Meanwhile, the end sections on Practical Issues and Moving Toward Speech to Speech Translation will allow us to discuss issues in incremental Speech Recognition and Text-to-Speech that are otherwise under-represented at a typical *ACL conference.

At most 33% are work by the presenters, and at least 77% are work by other researchers.

5 Diversity

Simultaneous translation techniques can greatly improve the efficiency of human communication across linguistic barriers. With this technology, you will be able to understand any foreign language by pulling out your smart phone to listen to the machine-generated simultaneous translation in your own language, with only less than 3 seconds delay. If you travel to a remote country, you will also be able to "talk" to the locals with this technology using your smart phone and headsets.

Both Mingbo Ma and Naveen Arivazhagan are junior instructors. Colin Cherry works at Google in Montreal, Liang Huang works Oregon State University in Corvallis, and Zhongjun He works at Baidu in Beijing.

6 Prerequisites

- Machine Learning: understand the basics of the sequence-to-sequence framework.

- Linguistics: understand basic syntactic structures and appreciate the vast amount of diversity of syntactic structures (esp. word order) among human languages

7 Small Reading List

Only the last two (33%) were co-authored by the presenters.

- Alvin Grissom II, He He, Jordan Boyd-Graber, John Morgan, and Hal Daumé III, Don't Until the Final Verb Wait: Reinforcement Learning for Simultaneous Machine Translation, EMNLP 2014.

When source and target language have drastically word orders difference, e.g., from verb-final languages (German) to verb-medial languages (English), the final inal verb is predicted in advance on source side to avoid long latency.

- He He, Alvin Grissom II, Jordan Boyd-Graber and Hal Daumé III, Syntax-based Rewriting for Simultaneous Machine Translation, EMNLP 2015.

A sentence rewriting method is proposed to generates more monotonic translations to improve the speed-accuracy tradeof. Several grammaticality and meaning-preserving syntactic transformation rules are applied to paraphrase reference translations to make their word order closer to the source language word order.

- Kyunghyun Cho and Masha Esipova, Can neural machine translation do simultaneous translation?, arXiv:1606.02012, 2016.

Several waiting criteria are manually designed to serve as translation polices to decide wait or read.

- Jiatao Gu, Graham Neubig, Kyunghyun Cho and Victor O.K. Li, Learning to Translate in Real-time with Neural Machine Translation, EACL 2017.

The authors proposed a NMT framework for simultaneous translation with a agent which learn to make decisions on when to translate or wait by interacting with a pre-trained NMT environment.

- Mingbo Ma, Liang Huang, Hao Xiong, Renjie Zheng, Kaibo Liu, Baigong Zheng, Chuanqiang Zhang, Zhongjun He, Hairong Liu, Xing Li, Hua Wu and Haifeng Wang, STACL: Simultaneous Translation with Implicit Anticipation and Controllable Latency using Prefix-to-Prefix Framework, ACL 2019.

Prefix-to-prefix framework is proposed for simultaneous translation which implicitly learns to anticipate in a single translation model. Within this framework, "wait-k" policy is trained to generate the target sentence simultaneously with the source sentence with k word delay.

- Naveen Arivazhagan, Colin Cherry, Wolfgang Macherey, Chung-Cheng Chiu, Semih Yavuz, Ruoming Pang Wei Li and Colin Raffel, Monotonic Infinite Lookback Attention for Simultaneous Machine Translation, ACL 2019.

  A Monotonic Infinite Lookback (MILk) technique is proposed to maintain both a hard, monotonic attention head to schedule the reading of the source sentence, and a soft attention head to extend from the monotonic head back to the beginning of the source. MILk is trained to learn a adaptive schedule by balancing the latency-quality trade-offs.

8 Presenters

- **Liang Huang** is an Assistant Professor at Oregon State University and a Distinguished Scientist of Baidu Research USA. He received a Best Paper Award at ACL 2008 and a Best Paper Honorable Mention at EMNLP 2016. He is an award-winning teacher and has given four (4) popular tutorials in COLING 2008, NAACL 2009, ACL 2014, and ACL 2015. He gave an invited talk at ACL 2019 on simultaneous translation.

- **Colin Cherry** is a research scientist at Google. He currently serves as secretary of NAACL and an action editor of TACL. He received Best Paper Award at NAACL 2009. He co-organized two workshops on deep learning for low-resource languages: DeepLo 2018 (at ACL 2018) and DeepLo 2019 (at EMNLP 2019). He also served as research program co-chair for AMTA 2018.

- **Mingbo Ma** is a Senior Research Scientist at Baidu Research USA. He received his Ph.D. from Oregon State University. He is a leading expert in simultaneous translation, and has published 4 papers on this topic.

- **Naveen Arizhabagan** is a Software Engineer at Google. He received BS from UIUC and MS from Stanford. He works on simultaneous translation, speech translation, zero-shot translation, and multilingual translation.

- **Zhongjun He** is a Distinguished Architect of Baidu Inc. He leads Baidu machine translation team and has released several versions of Baidu's simultaneous translation system since 2017. He organized the first simultaneous translation evaluation campaign in China in 2019 and released the Baidu Speech Translation Corpus.

9 Estimated Audience Size

150–200.

10 Special Technical Requirements

Internet access

11 Venue Preference

- First Choice: ACL

- Second Choice: EMNLP

12 Open Access

All materials (slides, videos, etc.) will be openly available online.