

ACL 2020

**Workshop on Automatic Simultaneous Translation
Challenges, Recent Advances, and Future Directions**

Proceedings of the Workshop

July, 10, 2020

©2020 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-952148-23-1 (Volume 1)

Introduction

Welcome to the First Workshop on Automatic Simultaneous Translation (AutoSimTrans). 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. It combines the AI technologies of machine translation (MT), automatic speech recognition (ASR), and text-to-speech synthesis (TTS), and is becoming a cutting-edge research field.

As an emerging and interdisciplinary field, simultaneous translation faces many great challenges, and is considered one of the holy grails of AI. This workshop will bring together researchers and practitioners in machine translation, speech processing, and human interpretation, to discuss recent advances and open challenges of simultaneous translation.

We organized a simultaneous translation shared task on Chinese-English. We released a dataset for open research, which covers speeches in a wide range of domains, such as IT, economy, culture, biology, arts, etc.

We also have two sets of keynote speakers: Hua Wu, Colin Cherry, Jordan Boyd-Graber, Qun Liu from simultaneous translation, Kay-Fan Cheung and Barry Slaughter Olsen from human interpretation research. We hope this workshop will greatly increase the communication and crossfertilization between the two fields.

We look forward to an exciting workshop.

Hua Wu, Colin Cherry, Liang Huang, Zhongjun He, Mark Liberman, James Cross, Yang Liu

Organizers:

Hua Wu, Baidu Inc.
Colin Cherry, Google
Liang Huang, Oregon State University and Baidu Research
Zhongjun He, Baidu Inc.
Mark Liberman, University of Pennsylvania
James Cross, Facebook
Yang Liu, Tsinghua University

Program Committee:

Mingbo Ma, Baidu Research, USA
Naveen Arivazhagan, Google, USA
Chung-Cheng Chiu, Google, USA
Kenneth Church, Baidu Research, USA
Yang Feng, CAS/ICT, China
George Foster, Google, Canada
Alvin Grissom II, Ursinus College, USA
He He, NYU, USA
Alina Karakanta, FBK-Trento, Italy
Wei Li, Google, USA
Hairong Liu, Baidu Research, USA
Kaibo Liu, Baidu Research, USA
Wolfgang Macherey, Google, USA
Jan Niehues, Maastricht U., Netherlands
Yusuke Oda, Google, Japan
Colin Raffel, Google, USA
Elizabeth Salesky, CMU, USA
Jiajun Zhang, CAS/IA, China
Ruiqing Zhang, Baidu Inc. China
Renjie Zheng, Oregon State Univ., USA

Invited Speakers:

Hua Wu, Chief Scientist of NLP, Baidu Inc., China
Colin Cherry, Research Scientist in Google Translate, Google Inc., Montreal, Canada
Jordan Boyd-Graber, Associate Professor, University of Maryland, USA
Qun Liu, Chief Scientist of Speech and Language Computing, Huawei Noah's Ark Lab, China
Kay-Fan Cheung, Associate Professor, The Hong Kong Polytechnic University,
Member of the International Association of Conference Interpreters (AIIC), China
Barry Slaughter Olsen, Professor, the Middlebury Institute of International Studies
and Conference Interpreter, Member of AIIC, USA

Table of Contents

<i>Dynamic Sentence Boundary Detection for Simultaneous Translation</i> Ruiqing Zhang and Chuanqiang Zhang	1
<i>End-to-End Speech Translation with Adversarial Training</i> Xuancai Li, Chen Kehai, Tiejun Zhao and Muyun Yang	10
<i>Robust Neural Machine Translation with ASR Errors</i> Haiyang Xue, Yang Feng, Shuhao Gu and Wei Chen	15
<i>Improving Autoregressive NMT with Non-Autoregressive Model</i> Long Zhou, Jiajun Zhang and Chengqing Zong	24
<i>Modeling Discourse Structure for Document-level Neural Machine Translation</i> Junxuan Chen, Xiang Li, Jiarui Zhang, Chulun Zhou, Jianwei Cui, Bin Wang and Jinsong Su ...	30
<i>BIT's system for the AutoSimTrans 2020</i> Minqin Li, Haodong Cheng, Yuanjie Wang, Sijia Zhang, Liting Wu and Yuhang Guo	37

Conference Program

Friday, July 10, 2020

8:50–9:00 *Opening Remarks*

9:00–11:00 Session 1

9:00–9:30 *Invited Talk 1: Colin Cherry*

9:30–10:00 *Invited Talk 2: Barry Slaughter Olsen*

10:00–10:30 *Invited Talk 3: Jordan Boyd-Graber*

10:30–11:00 *Q&A*

11:00–14:00 Lunch

17:00–19:00 Session 2

17:00–17:30 *Invited Talk 4: Hua Wu*

17:30–18:00 *Invited Talk 5: Kay-Fan Cheung*

18:00–18:30 *Invited Talk 6: Qun Liu*

18:30–19:00 *Q&A*

Friday, July 10, 2020 (continued)

19:00–19:30 Break

19:30–21:00 Session 3: Research Paper and System Description

19:30–19:40 *Dynamic Sentence Boundary Detection for Simultaneous Translation*
Ruiqing Zhang and Chuanqiang Zhang

19:40–19:50 *End-to-End Speech Translation with Adversarial Training*
Xuancai Li, Chen Kehai, Tiejun Zhao and Muyun Yang

19:50–20:00 *Robust Neural Machine Translation with ASR Errors*
Haiyang Xue, Yang Feng, Shuhao Gu and Wei Chen

20:00–20:10 *Improving Autoregressive NMT with Non-Autoregressive Model*
Long Zhou, Jiajun Zhang and Chengqing Zong

20:10–20:20 *Modeling Discourse Structure for Document-level Neural Machine Translation*
Junxuan Chen, Xiang Li, Jiarui Zhang, Chulun Zhou, Jianwei Cui, Bin Wang and Jinsong Su

20:20–20:30 *BIT's system for the AutoSimTrans 2020*
Minqin Li, Haodong Cheng, Yuanjie Wang, Sijia Zhang, Liting Wu and Yuhang Guo

20:30–21:00 Q&A

21:00–21:10 *Closing Remarks*