

WAT 2021

The 8th Workshop on Asian Translation

Proceedings of the Workshop

August 5–6, 2021
Bangkok, Thailand (online)

©2021 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

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 978-1-954085-63-3

Preface

Many Asian countries are rapidly growing these days and the importance of communicating and exchanging the information with these countries has intensified. To satisfy the demand for communication among these countries, machine translation technology is essential.

Machine translation technology has rapidly evolved recently and it is seeing practical use especially between European languages. However, the translation quality of Asian languages is not that high compared to that of European languages, and machine translation technology for these languages has not reached a stage of proliferation yet. This is not only due to the lack of the language resources for Asian languages but also due to the lack of techniques to correctly transfer the meaning of sentences from/to Asian languages. Consequently, a place for gathering and sharing the resources and knowledge about Asian language translation is necessary to enhance machine translation research for Asian languages.

The Conference on Machine Translation (WMT), the world's largest machine translation workshop, mainly targets on European language. The International Workshop on Spoken Language Translation (IWSLT) has spoken language translation tasks for some Asian languages using TED talk data, but there is no task for written language. The Workshop on Asian Translation (WAT) is an open machine translation evaluation campaign focusing on Asian languages. WAT gathers and shares the resources and knowledge of Asian language translation to understand the problems to be solved for the practical use of machine translation technologies among all Asian countries. WAT is unique in that it is an "open innovation platform": the test data is fixed and open, so participants can repeat evaluations on the same data and confirm changes in translation accuracy over time. WAT has no deadline for the automatic translation quality evaluation (continuous evaluation), so participants can submit translation results at any time.

Following the success of the previous WAT workshops (WAT2014 – WAT2020), WAT2021 will bring together machine translation researchers and users to try, evaluate, share and discuss brand-new ideas about machine translation. For the 8th WAT, we included several new translation tasks including Malayalam Visual Genome Task, MultiIndicMT, Restricted Translation Task and Ambiguous MSCOCO Task. We had 28 teams participated in the shared tasks and 24 teams submitted their translation results for the human evaluation. About 2,100 translation results were submitted to the automatic evaluation server, and selected submissions were manually evaluated. In addition to the shared tasks, WAT2021 also features research papers on topics related to machine translation, especially for Asian languages. The program committee accepted 5 research papers.

We are grateful to "SunFlare Co., Ltd.", "Kawamura International" and "Asia-Pacific Association for Machine Translation (AAMT)" for partially sponsoring the workshop. We would like to thank all the authors who submitted papers. We express our deepest gratitude to the committee members for their timely reviews. We also thank the ACL-IJCNLP 2021 organizers for their help with administrative matters.

WAT 2021 Organizers

Organizing Committee:

Toshiaki Nakazawa, The University of Tokyo, Japan
Hideki Nakayama, The University of Tokyo, Japan
Isao Goto, Japan Broadcasting Corporation (NHK), Japan
Hideya Mino, Japan Broadcasting Corporation (NHK), Japan
Chenchen Ding, National Institute of Information and Communications Technology (NICT), Japan
Raj Dabre, National Institute of Information and Communications Technology (NICT), Japan
Anoop Kunchukuttan, Microsoft AI and Research, India
Shohei Higashiyama, National Institute of Information and Communications Technology (NICT), Japan
Hiroshi Manabe, National Institute of Information and Communications Technology (NICT), Japan
Win Pa Pa, University of Computer Studies, Yangon (UCSY), Myanmar
Shantipriya Parida, Idiap Research Institute, Martigny, Switzerland
Ondřej Bojar, Charles University, Prague, Czech Republic
Chenhui Chu, Kyoto University, Japan
Akiko Eriguchi, Microsoft, USA
Kaori Abe, Tohoku University, Japan
Yusuke Oda, LegalForce, Japan
Katsuhito Sudoh, Nara Institute of Science and Technology (NAIST), Japan
Sadao Kurohashi, Kyoto University, Japan
Pushpak Bhattacharyya, Indian Institute of Technology Bombay (IIT), India

Program Committee:

Hailong Cao, Harbin Institute of Technology, China
Michael Carl, Kent State University, USA
Chenhui Chu, Kyoto University, Japan
Fabien Cromières, Free, France
Kenji Imamura, NICT, Japan
Yang Liu, Tsinghua University, China
Takashi Ninomiya, Ehime University, Japan
Masao Utiyama, NICT, Japan
Jiajun Zhang, Chinese Academy of Sciences, China

Technical Collaborators:

Luis Fernando D’Haro, Universidad Politécnica de Madrid, Spain

Rafael E. Banchs, Nanyang Technological University, Singapore

Haizhou Li, National University of Singapore, Singapore

Chen Zhang, National University of Singapore, Singapore

Table of Contents

Overview of the 8th Workshop on Asian Translation

Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Shohei Higashiyama, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar, Chenhui Chu, Akiko Eriguchi, Kaori Abe, Yusuke Oda and Sadao Kurohashi 1

NHK's Lexically-Constrained Neural Machine Translation at WAT 2021

Hideya Mino, Kazutaka Kinugawa, Hitoshi Ito, Isao Goto, Ichiro Yamada and Takenobu Tokunaga 46

Input Augmentation Improves Constrained Beam Search for Neural Machine Translation: NTT at WAT 2021

Katsuki Chousa and Makoto Morishita 53

NICT's Neural Machine Translation Systems for the WAT21 Restricted Translation Task

Zuchao Li, Masao Utiyama, Eiichiro Sumita and Hai Zhao 62

Machine Translation with Pre-specified Target-side Words Using a Semi-autoregressive Model

Seiichiro Kondo, Aomi Koyama, Tomoshige Kiyuna, Toshio Hirasawa and Mamoru Komachi .. 68

NECTEC's Participation in WAT-2021

Zar Zar Hlaing, Ye Kyaw Thu, Thazin Myint Oo, Mya Ei San, Sasiporn Usanavasin, Ponrudee Netisopakul and Thepchai Supnithi 74

Hybrid Statistical Machine Translation for English-Myanmar: UTYCC Submission to WAT-2021

Ye Kyaw Thu, Thazin Myint Oo, Hlaing Myat Nwe, Khaing Zar Mon, Nang Aeindray Kyaw, Naing Linn Phyto, Nann Hwan Khun and Hnin Aye Thant 83

NICT-2 Translation System at WAT-2021: Applying a Pretrained Multilingual Encoder-Decoder Model to Low-resource Language Pairs

Kenji Imamura and Eiichiro Sumita 90

Rakuten's Participation in WAT 2021: Examining the Effectiveness of Pre-trained Models for Multilingual and Multimodal Machine Translation

Raymond Hendy Susanto, Dongzhe Wang, Sunil Yadav, Mausam Jain and Ohnmar Htun 96

BTS: Back TranScription for Speech-to-Text Post-Processor using Text-to-Speech-to-Text

chanjun park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo and Heuiseok Lim 106

Zero-pronoun Data Augmentation for Japanese-to-English Translation

Ryokan Ri, Toshiaki Nakazawa and Yoshimasa Tsuruoka 117

Evaluation Scheme of Focal Translation for Japanese Partially Amended Statutes

Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa and Katsuhiko Toyama 124

TMU NMT System with Japanese BART for the Patent task of WAT 2021

Hwichan Kim and Mamoru Komachi 133

System Description for Transperfect

Wiktor Stribizew, Fred Bane, José Conceição and Anna Zaretskaya 138

<i>Bering Lab's Submissions on WAT 2021 Shared Task</i>	
Heesoo Park and Dongjun Lee	141
<i>NLPHut's Participation at WAT2021</i>	
Shantipriya Parida, Subhedarshi Panda, Ketan Kotwal, Amulya Ratna Dash, Satya Ranjan Dash, Yashvardhan Sharma, Petr Motlicek and Ondřej Bojar	146
<i>Improved English to Hindi Multimodal Neural Machine Translation</i>	
Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Darsh Kaushik, Partha Pakray and Sivaji Bandyopadhyay	155
<i>IITP at WAT 2021: System description for English-Hindi Multimodal Translation Task</i>	
Baban Gain, Dibyanayan Bandyopadhyay and Asif Ekbal	161
<i>ViTA: Visual-Linguistic Translation by Aligning Object Tags</i>	
Kshitij Gupta, Devansh Gautam and Radhika Mamidi	166
<i>TMEKU System for the WAT2021 Multimodal Translation Task</i>	
Yuting Zhao, Mamoru Komachi, Tomoyuki Kajiwara and Chenhui Chu	174
<i>Optimal Word Segmentation for Neural Machine Translation into Dravidian Languages</i>	
Prajit Dhar, Arianna Bisazza and Gertjan van Noord	181
<i>Itihasa: A large-scale corpus for Sanskrit to English translation</i>	
Rahul Aralikkatte, Miryam de Lhoneux, Anoop Kunchukuttan and Anders Søgaard	191
<i>NICT-5's Submission To WAT 2021: MBART Pre-training And In-Domain Fine Tuning For Indic Languages</i>	
Raj Dabre and Abhisek Chakrabarty	198
<i>How far can we get with one GPU in 100 hours? CoAStaL at MultiIndicMT Shared Task</i>	
Rahul Aralikkatte, Héctor Ricardo Murrieta Bello, Daniel Hershcovich, Marcel Bollmann and Anders Søgaard	205
<i>IIT Hyderabad Submission To WAT 2021: Efficient Multilingual NMT systems for Indian languages</i>	
Sourav Kumar, Salil Aggarwal and Dipti Sharma	212
<i>Language Relatedness and Lexical Closeness can help Improve Multilingual NMT: IITBombay@MultiIndicNMT WAT2021</i>	
Jyotsana Khatri, Nikhil Saini and Pushpak Bhattacharyya	217
<i>Samsung R&D Institute Poland submission to WAT 2021 Indic Language Multilingual Task</i>	
Adam Dobrowolski, Marcin Szymański, Marcin Chochowski and Paweł Przybysz	224
<i>Multilingual Machine Translation Systems at WAT 2021: One-to-Many and Many-to-One Transformer based NMT</i>	
Shivam Mhaskar, Aditya Jain, Aakash Banerjee and Pushpak Bhattacharyya	233
<i>IITP-MT at WAT2021: Indic-English Multilingual Neural Machine Translation using Romanized Vocabulary</i>	
Ramakrishna Appicharla, Kamal Kumar Gupta, Asif Ekbal and Pushpak Bhattacharyya	238
<i>ANVITA Machine Translation System for WAT 2021 MultiIndicMT Shared Task</i>	
Pavanpankaj Vegi, Sivabhavani J, Biswajit Paul, Chitra Viswanathan and Prasanna Kumar K R	244

Workshop Program

August 5-6, 2021 [UTC+0]

22:00–22:05 Welcome

Overview of the 8th Workshop on Asian Translation

Toshiaki Nakazawa, Hideki Nakayama, Chenchen Ding, Raj Dabre, Shohei Higashiyama, Hideya Mino, Isao Goto, Win Pa Pa, Anoop Kunchukuttan, Shantipriya Parida, Ondřej Bojar, Chenhui Chu, Akiko Eriguchi, Kaori Abe, Yusuke Oda and Sadao Kurohashi

22:05–22:45 Invited Talk I

Francisco Guzmán

22:45–22:50 Short Break

22:50–0:30 Shared Task I

Task Descriptions and Results (Restricted / ALT+UCSY / NICT-SAP)

Akiko Eriguchi, Chenchen Ding, and Raj Dabre

NHK's Lexically-Constrained Neural Machine Translation at WAT 2021

Hideya Mino, Kazutaka Kinugawa, Hitoshi Ito, Isao Goto, Ichiro Yamada and Takenobu Tokunaga

Input Augmentation Improves Constrained Beam Search for Neural Machine Translation: NTT at WAT 2021

Katsuki Chousa and Makoto Morishita

NICT's Neural Machine Translation Systems for the WAT21 Restricted Translation Task

Zuchao Li, Masao Utiyama, Eiichiro Sumita and Hai Zhao

Machine Translation with Pre-specified Target-side Words Using a Semi-autoregressive Model

Seiichiro Kondo, Aomi Koyama, Tomoshige Kiyuna, Tosho Hirasawa and Mamoru Komachi

NECTEC's Participation in WAT-2021

Zar Zar Hlaing, Ye Kyaw Thu, Thazin Myint Oo, Mya Ei San, Sasiporn Usanavasin, Ponrudee Netisopakul and Thepchai Supnithi

August 5-6, 2021 [UTC+0] (continued)

Hybrid Statistical Machine Translation for English-Myanmar: UTYCC Submission to WAT-2021

Ye Kyaw Thu, Thazin Myint Oo, Hlaing Myat Nwe, Khaing Zar Mon, Nang Aein-dray Kyaw, Naing Linn Phyto, Nann Hwan Khun and Hnin Aye Thant

NICT-2 Translation System at WAT-2021: Applying a Pretrained Multilingual Encoder-Decoder Model to Low-resource Language Pairs

Kenji Imamura and Eiichiro Sumita

Rakuten's Participation in WAT 2021: Examining the Effectiveness of Pre-trained Models for Multilingual and Multimodal Machine Translation

Raymond Hendy Susanto, Dongzhe Wang, Sunil Yadav, Mausam Jain and Ohnmar Htun

0:30–0:35 Short Break

0:35–1:15 Invited Talk II

Graham Neubig

1:15–1:20 Short Break

1:20–2:40 Research Paper / Findings I

BTS: Back TranScription for Speech-to-Text Post-Processor using Text-to-Speech-to-Text

chanjun park, Jaehyung Seo, Seolhwa Lee, Chanhee Lee, Hyeonseok Moon, Sugyeong Eo and Heuiseok Lim

Zero-pronoun Data Augmentation for Japanese-to-English Translation

Ryokan Ri, Toshiaki Nakazawa and Yoshimasa Tsuruoka

Evaluation Scheme of Focal Translation for Japanese Partially Amended Statutes

Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa and Katsuhiko Toyama

[ACL2021 Findings] Joint Optimization of Tokenization and Downstream Model

Tatsuya Hiraoka, Sho Takase, Kei Uchiumi, Atsushi Keyaki, and Naoaki Okazaki

August 5-6, 2021 [UTC+0] (continued)

2:40–4:00 Lunch Break (11:40-13:00 UTC+9)

4:00–5:40 Shared Task II

Task Descriptions and Results (JPC / Hindi&Malayalam Multimodal / Japanese Multimodal)

Shohei Higashiyama, Shantipriya Parida, Hideki Nakayama, and Chenhui Chu

TMU NMT System with Japanese BART for the Patent task of WAT 2021

Hwichan Kim and Mamoru Komachi

System Description for Transperfect

Wiktor Stribizew, Fred Bane, José Conceição and Anna Zaretskaya

Bering Lab's Submissions on WAT 2021 Shared Task

Heesoo Park and Dongjun Lee

NLPHut's Participation at WAT2021

Shantipriya Parida, Subhadarshi Panda, Ketan Kotwal, Amulya Ratna Dash, Satya Ranjan Dash, Yashvardhan Sharma, Petr Motlicek and Ondřej Bojar

Improved English to Hindi Multimodal Neural Machine Translation

Sahinur Rahman Laskar, Abdullah Faiz Ur Rahman Khilji, Darsh Kaushik, Partha Pakray and Sivaji Bandyopadhyay

IITP at WAT 2021: System description for English-Hindi Multimodal Translation Task

Baban Gain, Dibyanayan Bandyopadhyay and Asif Ekbal

ViTA: Visual-Linguistic Translation by Aligning Object Tags

Kshitij Gupta, Devansh Gautam and Radhika Mamidi

TMEKU System for the WAT2021 Multimodal Translation Task

Yuting Zhao, Mamoru Komachi, Tomoyuki Kajiwara and Chenhui Chu

August 5-6, 2021 [UTC+0] (continued)

5:40–4:45 Short Break

5:45–6:45 Research Paper / Findings II

Optimal Word Segmentation for Neural Machine Translation into Dravidian Languages

Prajit Dhar, Arianna Bisazza and Gertjan van Noord

Itihasa: A large-scale corpus for Sanskrit to English translation

Rahul Aralikkatte, Miryam de Lhoneux, Anoop Kunchukuttan and Anders Søgaard

[ACL2021 Findings] IndoCollex: A Testbed for Morphological Transformation of Indonesian Word Colloquialism

Haryo Akbarianto Wibowo, Made Nindyatama Nityasya, Afra Feyza Akyüek, Suci Fitriany, Alham Fikri Aji, Radityo Eko Prasoj, and Derry Tanti Wijaya

6:45–6:50 Short Break

6:50–8:20 Shared Task III

Task Descriptions and Results (MultiIndicMT)

Shantipriya Parida

NICT-5's Submission To WAT 2021: MBART Pre-training And In-Domain Fine Tuning For Indic Languages

Raj Dabre and Abhisek Chakrabarty

How far can we get with one GPU in 100 hours? CoAStAL at MultiIndicMT Shared Task

Rahul Aralikkatte, Héctor Ricardo Murrieta Bello, Daniel Hershcovich, Marcel Bollmann and Anders Søgaard

IIT Hyderabad Submission To WAT 2021: Efficient Multilingual NMT systems for Indian languages

Sourav Kumar, Salil Aggarwal and Dipti Sharma

Language Relatedness and Lexical Closeness can help Improve Multilingual NMT: IITBombay@MultiIndicNMT WAT2021

Jyotsana Khatri, Nikhil Saini and Pushpak Bhattacharyya

Samsung R&D Institute Poland submission to WAT 2021 Indic Language Multilingual Task

Adam Dobrowolski, Marcin Szymański, Marcin Chochowski and Paweł Przybyś

August 5-6, 2021 [UTC+0] (continued)

Multilingual Machine Translation Systems at WAT 2021: One-to-Many and Many-to-One Transformer based NMT

Shivam Mhaskar, Aditya Jain, Aakash Banerjee and Pushpak Bhattacharyya

IITP-MT at WAT2021: Indic-English Multilingual Neural Machine Translation using Romanized Vocabulary

Ramakrishna Appicharla, Kamal Kumar Gupta, Asif Ekbal and Pushpak Bhattacharyya

ANVITA Machine Translation System for WAT 2021 MultiIndicMT Shared Task

Pavanpankaj Vegi, Sivabhavani J, Biswajit Paul, Chitra Viswanathan and Prasanna Kumar K R

8:20–8:25 Closing

