

Eighth Conference on Machine Translation (WMT 2023)

Singapore
6 - 7 December 2023

Volume 1 of 2

ISBN: 978-1-7138-8587-0

Printed from e-media with permission by:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571



Some format issues inherent in the e-media version may also appear in this print version.

Copyright© (2023) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics
at the address below.

Association for Computational Linguistics
209 N. Eighth Street
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006

Fax: 1-570-476-0860

acl@aclweb.org

Additional copies of this publication are available from:

Curran Associates, Inc.
57 Morehouse Lane
Red Hook, NY 12571 USA
Phone: 845-758-0400
Fax: 845-758-2633
Email: curran@proceedings.com
Web: www.proceedings.com

Table of Contents

Findings of the 2023 Conference on Machine Translation (WMT23): LLMs Are Here but Not Quite There Yet

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Philipp Koehn, Benjamin Marie, Christof Monz, Makoto Morishita, Kenton Murray, Makoto Nagata, Toshiaki Nakazawa, Martin Popel, Maja Popović and Mariya Shmatova 1

Findings of the WMT 2023 Biomedical Translation Shared Task: Evaluation of ChatGPT 3.5 as a Comparison System

Mariana Neves, Antonio Jimeno Yepes, Aurélie Névéol, Rachel Bawden, Giorgio Maria Di Nunzio, Roland Roller, Philippe Thomas, Federica Vezzani, Maika Vicente Navarro, Lana Yeganova, Dina Wiemann and Cristian Grozea 43

Findings of the WMT 2023 Shared Task on Discourse-Level Literary Translation: A Fresh Orb in the Cosmos of LLMs

Longyue Wang, Zhaopeng Tu, Yan Gu, Siyou Liu, Dian Yu, Qingsong Ma, Chenyang Lyu, Liting Zhou, Chao-Hong Liu, Yufeng Ma, Weiyu Chen, Yvette Graham, Bonnie Webber, Philipp Koehn, Andy Way, Yulin Yuan and Shuming Shi 55

Findings of the Second WMT Shared Task on Sign Language Translation (WMT-SLT23)

Mathias Müller, Malihe Alikhani, Eleftherios Avramidis, Richard Bowden, Annelies Braffort, Necati Cihan Camgöz, Sarah Ebling, Cristina España-Bonet, Anne Göhring, Roman Grundkiewicz, Mert Inan, Zifan Jiang, Oscar Koller, Amit Moryossef, Annette Rios, Dimitar Shterionov, Sandra Sidler-Miserez, Katja Tissi and Davy Van Landuyt 68

Findings of the WMT 2023 Shared Task on Parallel Data Curation

Steve Sloto, Brian Thompson, Huda Khayrallah, Tobias Domhan, Thamme Gowda and Philipp Koehn 95

Samsung R&D Institute Philippines at WMT 2023

Jan Christian Blaise Cruz 103

NAIST-NICT WMT'23 General MT Task Submission

Hiroyuki Deguchi, Kenji Imamura, Yuto Nishida, Yusuke Sakai, Justin Vasselli and Taro Watanabe 110

CUNI at WMT23 General Translation Task: MT and a Genetic Algorithm

Josef Jon, Martin Popel and Ondřej Bojar 119

SKIM at WMT 2023 General Translation Task

Keito Kudo, Takumi Ito, Makoto Morishita and Jun Suzuki 128

KYB General Machine Translation Systems for WMT23

Ben LI, Yoko Matsuzaki and Shivam Kalkar 137

Yishu: Yishu at WMT2023 Translation Task

Luo Min, yixin tan and Qiulin Chen 143

PROMT Systems for WMT23 Shared General Translation Task

Alexander Molchanov and Vladislav Kovalenko 150

<i>AIST AIRC Submissions to the WMT23 Shared Task</i>	
Matiss Rikters and Makoto Miwa	155
<i>MUNI-NLP Submission for Czech-Ukrainian Translation Task at WMT23</i>	
Pavel Rychly and Yuliia Teslia	162
<i>Exploring Prompt Engineering with GPT Language Models for Document-Level Machine Translation: Insights and Findings</i>	
Yangjian Wu and Gang Hu	166
<i>Treating General MT Shared Task as a Multi-Domain Adaptation Problem: HW-TSC's Submission to the WMT23 General MT Shared Task</i>	
Zhanglin Wu, Daimeng Wei, Zongyao Li, Zhengzhe YU, Shaojun Li, Xiaoyu Chen, Hengchao Shang, Jiabin GUO, Yuhao Xie, Lizhi Lei, Hao Yang and Yanfei Jiang	170
<i>UvA-MT's Participation in the WMT 2023 General Translation Shared Task</i>	
Di Wu, Shaomu Tan, David Stap, Ali Araabi and Christof Monz	175
<i>Achieving State-of-the-Art Multilingual Translation Model with Minimal Data and Parameters</i>	
Hui Zeng	181
<i>IOL Research Machine Translation Systems for WMT23 General Machine Translation Shared Task</i>	
Wenbo Zhang	187
<i>GTCOM and DLUT's Neural Machine Translation Systems for WMT23</i>	
Hao Zong	192
<i>RoCS-MT: Robustness Challenge Set for Machine Translation</i>	
Rachel Bawden and Benoît Sagot	198
<i>Multifaceted Challenge Set for Evaluating Machine Translation Performance</i>	
Xiaoyu Chen, Daimeng Wei, Zhanglin Wu, Ting Zhu, Hengchao Shang, Zongyao Li, Jiabin GUO, Ning Xie, Lizhi Lei, Hao Yang and Yanfei Jiang	217
<i>Linguistically Motivated Evaluation of the 2023 State-of-the-art Machine Translation: Can ChatGPT Outperform NMT?</i>	
Shushen Manakhimova, Eleftherios Avramidis, Vivien Macketanz, Ekaterina Lapshinova-Koltunski, Sergei Bagdasarov and Sebastian Möller	224
<i>IIIT HYD's Submission for WMT23 Test-suite Task</i>	
Ananya Mukherjee and Manish Shrivastava	246
<i>Test Suites Task: Evaluation of Gender Fairness in MT with MuST-SHE and INES</i>	
Beatrice Savoldi, Marco Gaido, Matteo Negri and Luisa Bentivogli	252
<i>Biomedical Parallel Sentence Retrieval Using Large Language Models</i>	
Sheema Firdous and Sadaf Abdul Rauf	263
<i>The Path to Continuous Domain Adaptation Improvements by HW-TSC for the WMT23 Biomedical Translation Shared Task</i>	
Zhanglin Wu, Daimeng Wei, Zongyao Li, Zhengzhe YU, Shaojun Li, Xiaoyu Chen, Hengchao Shang, Jiabin GUO, Yuhao Xie, Lizhi Lei, Hao Yang and Yanfei Jiang	271

<i>Investigating Techniques for a Deeper Understanding of Neural Machine Translation (NMT) Systems through Data Filtering and Fine-tuning Strategies</i>	
Lichao Zhu, Maria Zimina, Maud Bénard, Behnoosh Namdar, Nicolas Ballier, Guillaume Wisniewski and Jean-Baptiste Yunès	275
<i>MAX-ISI System at WMT23 Discourse-Level Literary Translation Task</i>	
Li An, Linghao Jin and Xuezhe Ma	282
<i>The MAKE-NMTVIZ System Description for the WMT23 Literary Task</i>	
Fabien Lopez, Gabriela González, Damien Hansen, Mariam Nakhle, Behnoosh Namdarzadeh, Nicolas Ballier, Marco Dinarelli, Emmanuelle Esperança-Rodier, Sui He, Sadaf Mohseni, Caroline Rossi, Didier Schwab, Jun Yang, Jean-Baptiste Yunès and Lichao Zhu	287
<i>DUTNLP System for the WMT2023 Discourse-Level Literary Translation</i>	
Anqi Zhao, Kaiyu Huang, Hao Yu and Degen Huang	296
<i>HW-TSC's Submissions to the WMT23 Discourse-Level Literary Translation Shared Task</i>	
Yuhao Xie, Zongyao Li, Zhanglin Wu, Daimeng Wei, Xiaoyu Chen, Zhiqiang Rao, Shaojun Li, Hengchao Shang, Jiaxin GUO, Lizhi Lei, Hao Yang and Yanfei Jiang	302
<i>TJUNLP: System Description for the WMT23 Literary Task in Chinese to English Translation Direction</i>	
Shaolin Zhu and Deyi Xiong	307
<i>Machine Translation for Nko: Tools, Corpora, and Baseline Results</i>	
Moussa Doumbouya, Baba Mamadi Diané, Solo Farabado Cissé, Djibrila Diané, Abdoulaye Sow, Séré Moussa Doumbouya, Daouda Bangoura, Fodé Moriba Bayo, Ibrahima Sory 2. Conde, Kalo Mory Diané, Chris Piech and Christopher Manning	312
<i>TTIC's Submission to WMT-SLT 23</i>	
Marcelo Sandoval-Castaneda, Yanhong Li, Bowen Shi, Diane Brentari, Karen Livescu and Gregory Shakhnarovich	344
<i>KnowComp Submission for WMT23 Sign Language Translation Task</i>	
Baixuan Xu, Haochen Shi, Tianshi Zheng, Qing Zong, Weiqi Wang, Zhaowei Wang and Yangqiu Song	351
<i>A Fast Method to Filter Noisy Parallel Data WMT2023 Shared Task on Parallel Data Curation</i>	
Nguyen-Hoang Minh-Cong, Nguyen Van Vinh and Nguyen Le-Minh	359
<i>A Sentence Alignment Approach to Document Alignment and Multi-faceted Filtering for Curating Parallel Sentence Pairs from Web-crawled Data</i>	
Steinthor Steingrímsson	366
<i>Document-Level Language Models for Machine Translation</i>	
Frithjof Petrick, Christian Herold, Pavel Petrushkov, Shahram Khadivi and Hermann Ney	375
<i>ChatGPT MT: Competitive for High- (but Not Low-) Resource Languages</i>	
Nathaniel Robinson, Perez Ogayo, David R. Mortensen and Graham Neubig	392
<i>Large Language Models Effectively Leverage Document-level Context for Literary Translation, but Critical Errors Persist</i>	
Marzena Karpinska and Mohit Iyyer	419
<i>Identifying Context-Dependent Translations for Evaluation Set Production</i>	
Rachel Wicks and Matt Post	452

<i>Machine Translation with Large Language Models: Prompting, Few-shot Learning, and Fine-tuning with QLoRA</i>	
Xuan Zhang, Navid Rajabi, Kevin Duh and Philipp Koehn	468
<i>Towards Effective Disambiguation for Machine Translation with Large Language Models</i>	
Vivek Iyer, Pinzhen Chen and Alexandra Birch	482
<i>A Closer Look at Transformer Attention for Multilingual Translation</i>	
Jingyi Zhang, Gerard de Melo, Hongfei Xu and Kehai Chen	496
<i>Bridging the Gap between Position-Based and Content-Based Self-Attention for Neural Machine Translation</i>	
Felix Schmidt and Mattia Di Gangi	507
<i>Visual Prediction Improves Zero-Shot Cross-Modal Machine Translation</i>	
Tosho Hirasawa, Emanuele Bugliarello, Desmond Elliott and Mamoru Komachi	522
<i>The Gender-GAP Pipeline: A Gender-Aware Polyglot Pipeline for Gender Characterisation in 55 Languages</i>	
Benjamin Muller, Belen Alastruey, Prangthip Hansanti, Elahe Kalbassi, Christophe Ropers, Eric Smith, Adina Williams, Luke Zettlemoyer, Pierre Andrews and Marta R. Costa-jussà	536
<i>Towards Better Evaluation for Formality-Controlled English-Japanese Machine Translation</i>	
Edison Marrese-Taylor, Pin Chen Wang and Yutaka Matsuo	551
<i>There's No Data like Better Data: Using QE Metrics for MT Data Filtering</i>	
Jan-Thorsten Peter, David Vilar, Daniel Deutsch, Mara Finkelstein, Juraj Juraska and Markus Freitag	561
<i>Results of WMT23 Metrics Shared Task: Metrics Might Be Guilty but References Are Not Innocent</i>	
Markus Freitag, Nitika Mathur, Chi-kiu Lo, Eleftherios Avramidis, Ricardo Rei, Brian Thompson, Tom Kocmi, Frederic Blain, Daniel Deutsch, Craig Stewart, Chrysoula Zerva, Sheila Castilho, Alon Lavie and George Foster	578
<i>Findings of the WMT 2023 Shared Task on Quality Estimation</i>	
Frederic Blain, Chrysoula Zerva, Ricardo Ribeiro, Nuno M. Guerreiro, Diptesh Kanojia, José G. C. de Souza, Beatriz Silva, Tânia Vaz, Yan Jingxuan, Fatemeh Azadi, Constantin Orasan and André Martins	629
<i>Findings of the Word-Level AutoCompletion Shared Task in WMT 2023</i>	
Lemao Liu, Francisco Casacuberta, George Foster, Guoping Huang, Philipp Koehn, Geza Kovacs, Shuming Shi, Taro Watanabe and Chengqing Zong	654
<i>Findings of the WMT 2023 Shared Task on Machine Translation with Terminologies</i>	
Kirill Semenov, Vilém Zouhar, Tom Kocmi, Dongdong Zhang, Wangchunshu Zhou and Yuchen Eleanor Jiang	663
<i>Findings of the WMT 2023 Shared Task on Automatic Post-Editing</i>	
Pushpak Bhattacharyya, Rajen Chatterjee, Markus Freitag, Diptesh Kanojia, Matteo Negri and Marco Turchi	672
<i>Findings of the WMT 2023 Shared Task on Low-Resource Indic Language Translation</i>	
Santanu Pal, Partha Pakray, Sahinur Rahman Laskar, Lenin Laitonjam, Vanlalmuansangi Khenglawt, Sunita Warjri, Pankaj Kundan Dadure and Sandeep Kumar Dash	682

<i>ACES: Translation Accuracy Challenge Sets at WMT 2023</i>	
Chantal Amrhein, Nikita Moghe and Liane Guillou	695
<i>Challenging the State-of-the-art Machine Translation Metrics from a Linguistic Perspective</i>	
Eleftherios Avramidis, Shushen Manakhimova, Vivien Macketanz and Sebastian Möller	713
<i>Tokengram_F, a Fast and Accurate Token-based chrF++ Derivative</i>	
Sören DREANO, Derek Molloy and Noel Murphy	730
<i>Embed_Llama: Using LLM Embeddings for the Metrics Shared Task</i>	
Sören DREANO, Derek Molloy and Noel Murphy	738
<i>eBLEU: Unexpectedly Good Machine Translation Evaluation Using Simple Word Embeddings</i>	
Muhammad Elnokrashy and Tom Kocmi	746
<i>Cometoid: Distilling Strong Reference-based Machine Translation Metrics into Even Stronger Quality Estimation Metrics</i>	
Thamme Gowda, Tom Kocmi and Marcin Junczys-Dowmunt	751
<i>MetricX-23: The Google Submission to the WMT 2023 Metrics Shared Task</i>	
Juraj Juraska, Mara Finkelstein, Daniel Deutsch, Aditya Siddhant, Mehdi Mirzazadeh and Markus Freitag	756
<i>GEMBA-MQM: Detecting Translation Quality Error Spans with GPT-4</i>	
Tom Kocmi and Christian Federmann	768
<i>Metric Score Landscape Challenge (MSLC23): Understanding Metrics' Performance on a Wider Landscape of Translation Quality</i>	
Chi-kiu Lo, Samuel Larkin and Rebecca Knowles	776
<i>MEE4 and XLSim : IIIT HYD's Submissions' for WMT23 Metrics Shared Task</i>	
Ananya Mukherjee and Manish Shrivastava	800
<i>Quality Estimation Using Minimum Bayes Risk</i>	
Subhajit Naskar, Daniel Deutsch and Markus Freitag	806
<i>Evaluating Metrics for Document-context Evaluation in Machine Translation</i>	
Vikas Raunak, Tom Kocmi and Matt Post	812
<i>Semantically-Informed Regressive Encoder Score</i>	
Vasiliy Viskov, George Kokush, Daniil Larionov, Steffen Eger and Alexander Panchenko	815
<i>Empowering a Metric with LLM-assisted Named Entity Annotation: HW-TSC's Submission to the WMT23 Metrics Shared Task</i>	
Zhanglin Wu, Yilun Liu, Min Zhang, Xiaofeng Zhao, Junhao Zhu, Ming Zhu, Xiaosong Qiao, Jingfei Zhang, Ma Miaomiao, Zhao Yanqing, Song Peng, shimin tao, Hao Yang and Yanfei Jiang ..	822
<i>Unify Word-level and Span-level Tasks: NJUNLP's Participation for the WMT2023 Quality Estimation Shared Task</i>	
Xiang Geng, Zhejian Lai, Yu Zhang, shimin tao, Hao Yang, Jiajun CHEN and Shujian Huang .	829
<i>HW-TSC 2023 Submission for the Quality Estimation Shared Task</i>	
Yuang Li, Chang Su, Ming Zhu, Mengyao Piao, Xinglin Lyu, Min Zhang and Hao Yang	835

<i>Scaling up CometKiwi: Unbabel-IST 2023 Submission for the Quality Estimation Shared Task</i>	
Ricardo Rei, Nuno M. Guerreiro, JosÃ© Pombal, Daan van Stigt, Marcos Treviso, Luisa Coheur, JosÃ© G. C. de Souza and AndrÃ© Martins	841
<i>SurreyAI 2023 Submission for the Quality Estimation Shared Task</i>	
Archchana Sindhujan, Diptesh Kanojia, Constantin Orasan and Tharindu Ranasinghe	849
<i>MMT's Submission for the WMT 2023 Quality Estimation Shared Task</i>	
Yulong Wu, Viktor Schlegel, Daniel Beck and Riza Batista-Navarro	856
<i>IOL Research's Submission for WMT 2023 Quality Estimation Shared Task</i>	
ZEYU YAN	863
<i>SJTU-MTLAB's Submission to the WMT23 Word-Level Auto Completion Task</i>	
Xingyu Chen and Rui Wang	872
<i>PRHLT's Submission to WLAC 2023</i>	
Angel Navarro, Miguel Domingo and Francisco Casacuberta	877
<i>KnowComp Submission for WMT23 Word-Level AutoCompletion Task</i>	
Yi Wu, Haochen Shi, Weiqi Wang and Yangqiu Song	882
<i>Terminology-Aware Translation with Constrained Decoding and Large Language Model Prompting</i>	
Nikolay Bogoychev and Pinzhen Chen	890
<i>Lingua Custodia's Participation at the WMT 2023 Terminology Shared Task</i>	
Jingshu Liu, Mariam Nakhel, Gaëtan Caillout and Raheel Qadar	897
<i>Domain Terminology Integration into Machine Translation: Leveraging Large Language Models</i>	
Yasmin Moslem, Gianfranco Romani, Mahdi Molaei, John D. Kelleher, Rejwanul Haque and Andy Way	902
<i>OPUS-CAT Terminology Systems for the WMT23 Terminology Shared Task</i>	
Tommi Nieminen	912
<i>VARCO-MT: NCSOFT's WMT'23 Terminology Shared Task Submission</i>	
Geon Woo Park, Junghwa Lee, Meiyong Ren, Allison Shindell and Yeonsoo Lee	919
<i>HW-TSC's Participation in the WMT 2023 Automatic Post Editing Shared Task</i>	
Jiawei Yu, Min Zhang, Zhao Yanqing, Xiaofeng Zhao, Yuang Li, Su Chang, Yinglu Li, Ma Miaomiao, shimin tao and Hao Yang	926
<i>Neural Machine Translation for English - Manipuri and English - Assamese</i>	
Goutam Agrawal, Rituraj Das, Anupam Biswas and Dalton Meitei Thounaojam	931
<i>GUIT-NLP's Submission to Shared Task: Low Resource Indic Language Translation</i>	
Mazida Ahmed, Kuwali Talukdar, Parvez Boruah, Prof. Shikhar Kumar Sarma and Kishore Kashyap	935
<i>NICT-AI4B's Submission to the Indic MT Shared Task in WMT 2023</i>	
Raj Dabre, Jay Gala and Pranjal Chitale	941
<i>Machine Translation Advancements for Low-Resource Indian Languages in WMT23: CFILT-IITB's Effort for Bridging the Gap</i>	
Pranav Gaikwad, Meet Doshi, Sourabh Deoghare and Pushpak Bhattacharyya	950

<i>Low-Resource Machine Translation Systems for Indic Languages</i> Ivana Kvapilíková and Ondřej Bojar	954
<i>MUNI-NLP Systems for Low-resource Indic Machine Translation</i> Edoardo Signoroni and Pavel Rychly	959
<i>NITS-CNLP Low-Resource Neural Machine Translation Systems of English-Manipuri Language Pair</i> Kshetrimayum Boynao Singh, Avichandra Singh Ningthoujam, Loitongbam Sanayai Meetei, Sivaji Bandyopadhyay and Thoudam Doren Singh	967
<i>IACS-LRILT: Machine Translation for Low-Resource Indic Languages</i> Dhairya Suman, Atanu Mandal, Santanu Pal and Sudip Naskar	972
<i>IOL Research Machine Translation Systems for WMT23 Low-Resource Indic Language Translation Shared Task</i> Wenbo Zhang	978
<i>Trained MT Metrics Learn to Cope with Machine-translated References</i> Jannis Vamvas, Tobias Domhan, Sony Trenous, Rico Sennrich and Eva Hasler	983
<i>Training and Meta-Evaluating Machine Translation Evaluation Metrics at the Paragraph Level</i> Daniel Deutsch, Juraj Juraska, Mara Finkelstein and Markus Freitag	996
<i>Automating Behavioral Testing in Machine Translation</i> Javier Ferrando, Matthias Sperber, Hendra Setiawan, Dominic Telaar and Saša Hasan	1014
<i>One Wide Feedforward Is All You Need</i> Telmo Pires, António Vilarinho Lopes, Yannick Assogba and Hendra Setiawan	1031
<i>A Benchmark for Evaluating Machine Translation Metrics on Dialects without Standard Orthography</i> Noëmi Aepli, Chantal Amrhein, Florian Schottnann and Rico Sennrich	1045
<i>The Devil Is in the Errors: Leveraging Large Language Models for Fine-grained Machine Translation Evaluation</i> Patrick Fernandes, Daniel Deutsch, Mara Finkelstein, Parker Riley, André Martins, Graham Neubig, Ankush Garg, Jonathan Clark, Markus Freitag and Orhan Firat	1066