

# **7th Workshop on Technologies for Machine Translation of Low Resource Languages (LoResMT 2024)**

Bangkok, Thailand  
15 August 2024

ISBN: 979-8-3313-0169-9

**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© (2024) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2025)

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](mailto: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](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Tuning LLMs with Contrastive Alignment Instructions for Machine Translation in Unseen, Low-resource Languages</i>	
Zhuoyuan Mao and Yen Yu .....	1
<i>HeSum: a Novel Dataset for Abstractive Text Summarization in Hebrew</i>	
Itai Mondshine, Tzuf Paz-Argaman, Asaf Achi Mordechai and Reut Tsarfaty .....	26
<i>KpopMT: Translation Dataset with Terminology for Kpop Fandom</i>	
JiWoo Kim, Yunsu Kim and JinYeong Bak .....	37
<i>Challenges in Urdu Machine Translation</i>	
Abdul Basit, Abdul Hameed Azeemi and Agha Ali Raza .....	44
<i>Linguistically Informed Transformers for Text to American Sign Language Translation</i>	
Abhishek Bharadwaj Varanasi, Manjira Sinha and Tirthankar Dasgupta .....	50
<i>Low-Resource Cross-Lingual Summarization through Few-Shot Learning with Large Language Models</i>	
Gyutae Park, Seojin Hwang and Hwanhee Lee .....	57
<i>Enhancing Low-Resource NMT with a Multilingual Encoder and Knowledge Distillation: A Case Study</i>	
Aniruddha Roy, Pretam Ray, Ayush Maheshwari, Sudeshna Sarkar and Pawan Goyal .....	64
<i>Leveraging Mandarin as a Pivot Language for Low-Resource Machine Translation between Cantonese and English</i>	
King Yiu Suen, Rudolf Chow and Albert Y.s. Lam .....	74
<i>Enhancing Turkish Word Segmentation: A Focus on Borrowed Words and Invalid Morpheme</i>	
Soheila Behrooznia, Ebrahim Ansari and Zdenek Zabokrtsky .....	85
<i>Super donors and super recipients: Studying cross-lingual transfer between high-resource and low-resource languages</i>	
Vitaly Protasov, Elisei Stakovskii, Ekaterina Voloshina, Tatiana Shavrina and Alexander Panchenko	94
<i>Tokenisation in Machine Translation Does Matter: The impact of different tokenisation approaches for Maltese</i>	
Kurt Abela, Kurt Micallef, Marc Tanti and Claudia Borg .....	109
<i>Machine Translation Through Cultural Texts: Can Verses and Prose Help Low-Resource Indigenous Models?</i>	
Antoine Cadotte, Nathalie André and Fatiha Sadat .....	121
<i>Rule-Based, Neural and LLM Back-Translation: Comparative Insights from a Variant of Ladin</i>	
Samuel Frontull and Georg Moser .....	128
<i>AGE: Amharic, Ge'ez and English Parallel Dataset</i>	
Henok Biadglign Ademtew and Mikiyas Girma Birbo .....	139
<i>Learning-From-Mistakes Prompting for Indigenous Language Translation</i>	
You Cheng Liao, Chen-Jui Yu, Chi-Yi Lin, He-Feng Yun, Yen-Hsiang Wang, Hsiao-Min Li and Yao-Chung Fan .....	146
<i>Adopting Ensemble Learning for Cross-lingual Classification of Crisis-related Text On Social Media</i>	
Shareefa Ahmed Al Amer, Mark G. Lee and Phillip Smith .....	159

<i>Finetuning End-to-End Models for Estonian Conversational Spoken Language Translation</i>	
Tiia Sildam, Andra Velve and Tanel Alumäe .....	166
<i>Benchmarking Low-Resource Machine Translation Systems</i>	
Ana Alexandra Morim Da Silva, Nikit Srivastava, Tatiana Moteu Ngoli, Michael Röder, Diego Moussallem and Axel-Cyrille Ngonga Ngomo .....	175
<i>Rosetta Balcanica: Deriving a "Gold Standard" Neural Machine Translation (NMT) Parallel Dataset from High-Fidelity Resources for Western Balkan Languages</i>	
Edmon Begoli, Maria Mahbub and Sudarshan Srinivasan .....	186
<i>Irish-based Large Language Model with Extreme Low-Resource Settings in Machine Translation</i>	
Khanh-Tung Tran, Barry O’Sullivan and Hoang D. Nguyen .....	193