

Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)

Albuquerque, New Mexico, USA
3 May 2025

ISBN: 979-8-3313-1923-6

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© (2025) 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

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

<i>Comparative Evaluation of Machine Translation Models Using Human-Translated Social Media Posts as References: Human-Translated Datasets</i>	
Shareefa Ahmed Al Amer, Mark G. Lee and Phillip Smith	1
<i>Enhanced Zero-Shot Machine Translation via Fixed Prefix Pair Bootstrapping</i>	
Van-Hien Tran and Masao Utiyama	10
<i>UTER: Capturing the Human Touch in Evaluating Morphologically Rich and Low-Resource Languages</i>	
Samy Ouzerrout	16
<i>From Text to Multi-Modal: Advancing Low-Resource-Language Translation through Synthetic Data Generation and Cross-Modal Alignments</i>	
Bushi Xiao, Qian Shen and Daisy Zhe Wang	24
<i>Wenzhou Dialect Speech to Mandarin Text Conversion</i>	
Zhipeng Gao, Akihiro Tamura and Tsuneo Kato	36
<i>Fostering Digital Inclusion for Low-Resource Nigerian Languages: A Case Study of Igbo and Nigerian Pidgin</i>	
Ebelechukwu Nwafor and Minh Phuc Nguyen	44
<i>Low-resource Machine Translation: what for? who for? An observational study on a dedicated Tetun language translation service</i>	
Raphael Merx, Adérito José Guterres Correia, Hanna Suominen and Ekaterina Vylomova	54
<i>Jamo-Level Subword Tokenization in Low-Resource Korean Machine Translation</i>	
Junyoung Lee, Marco Cognetta, Sangwhan Moon and Naoaki Okazaki	66
<i>Beyond English: The Impact of Prompt Translation Strategies across Languages and Tasks in Multilingual LLMs</i>	
Itai Mondshine, Tzuf Paz-Argaman and Reut Tsarfaty	81
<i>ModeLing: A Novel Dataset for Testing Linguistic Reasoning in Language Models</i>	
Nathan Andrew Chi, Teodor Malchev, Riley Kong, Ryan Andrew Chi, Lucas Huang, Ethan A Chi, R. Thomas McCoy and Dragomir Radev	105
<i>Multilingual State Space Models for Structured Question Answering in Indic Languages</i>	
Arpita Vats, Rahul Raja, Mrinal Mathur, Aman Chadha and Vinija Jain	115
<i>Parallel Corpora for Machine Translation in Low-Resource Indic Languages: A Comprehensive Review</i>	
Rahul Raja and Arpita Vats	129
<i>Low-Resource Transliteration for Roman-Urdu and Urdu Using Transformer-Based Models</i>	
Umer Butt, Stalin Varanasi and Günter Neumann	144
<i>Building Data Infrastructure for Low-Resource Languages</i>	
Sarah K. K. Luger, Rafael Mosquera and Pedro Ortiz Suarez	154
<i>Encoder-Aware Sequence-Level Knowledge Distillation for Low-Resource Neural Machine Translation</i>	
Menan Velayuthan, Nisansa De Silva and Surangika Ranathunga	161
<i>PahGen: Generating Ancient Pahlavi Text via Grammar-guided Zero-shot Translation</i>	
Farhan Farsi, Parnian Fazel, Farzaneh Goshtasb, Nadia Hajipour, Sadra Sabouri, Ehsaneddin Asgari and Hossein Sameti	171

<i>Limitations of Religious Data and the Importance of the Target Domain: Towards Machine Translation for Guinea-Bissau Creole</i>	
Jacqueline Rowe, Edward Gow-Smith and Mark Hepple	183