

# **Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2024)**

Languages of Latin America

Student Research Workshop

Mexico City, Mexico  
18 June 2024

ISBN: 978-1-7138-9898-6

**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© (2024) 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](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>Systematic Analysis for Pretrained Language Model Priming for Parameter-Efficient Fine-tuning</i> Shih-Cheng Huang, Shih-Heng Wang, Min-Han Shih, Saurav Sahay and Hung-yi Lee . . . . .	1
<i>Rephrasing Invokes Better Generations for Large Language Models</i> Haoran Yang, Hongyuan Lu and Wai Lam . . . . .	8
<i>Exploring Compositional Generalization of Large Language Models</i> Haoran Yang, Hongyuan Lu, Wai Lam and Deng Cai . . . . .	16
<i>Explainable CED: A Dataset for Explainable Critical Error Detection in Machine Translation</i> Dahyun Jung, Sugyeong Eo, Chanjun Park and Heuseok Lim . . . . .	25
<i>SMARTR: A Framework for Early Detection using Survival Analysis of Longitudinal Texts</i> Jean-Thomas Baillargeon and Luc Lamontagne . . . . .	36
<i>Fast Exact Retrieval for Nearest-neighbor Lookup (FERN)</i> Richard Zhu . . . . .	42
<i>Start Simple: Progressive Difficulty Multitask Learning</i> Yunfei Luo, Yuyang Liu, Rukai Cai and Tauhidur Rahman . . . . .	48
<i>LUCID: LLM-Generated Utterances for Complex and Interesting Dialogues</i> Joe Stacey, Jianpeng Cheng, John Torr, Tristan Guigue, Joris Driesen, Alexandru Coca, Mark Gaynor and Anders Johannsen . . . . .	56
<i>Fine-tuning Pre-trained Named Entity Recognition Models For Indian Languages</i> Sankalp Bahad, Pruthwik Mishra, Parameswari Krishnamurthy and Dipti Sharma . . . . .	75
<i>Knowledge-centered conversational agents with a drive to learn</i> Selene Baez Santamaria . . . . .	83
<i>Exploring Inherent Biases in LLMs within Korean Social Context: A Comparative Analysis of ChatGPT and GPT-4</i> Seungyoon Lee, Dong Jun Kim, Dahyun Jung, Chanjun Park and Heuseok Lim . . . . .	93
<i>To Clarify or not to Clarify: A Comparative Analysis of Clarification Classification with Fine-Tuning, Prompt Tuning, and Prompt Engineering</i> Alina Leippert, Tatiana Anikina, Bernd Kiefer and Josef Van Genabith . . . . .	105
<i>Detecting Response Generation Not Requiring Factual Judgment</i> Ryohei Kamei, Daiki Shiono, Reina Akama and Jun Suzuki . . . . .	116
<i>Unknown Script: Impact of Script on Cross-Lingual Transfer</i> Wondimagegnhue Tufa, Iliia Markov and Piek Vossen . . . . .	124
<i>Improving Repository-level Code Search with Text Conversion</i> Mizuki Kondo, Daisuke Kawahara and Toshiyuki Kurabayashi . . . . .	130
<i>Improving Multi-lingual Alignment Through Soft Contrastive Learning</i> Minsu Park, Seyeon Choi, Chanyeol Choi, Jun-Seong Kim and Jy-yong Sohn . . . . .	138
<i>Few-Shot Event Argument Extraction Based on a Meta-Learning Approach</i> Aboubacar Tuo, Romaric Besançon, Olivier Ferret and Julien Tourille . . . . .	146

<i>Investigating Web Corpus Filtering Methods for Language Model Development in Japanese</i> Rintaro Enomoto, Arseny Tolmachev, Takuro Niitsuma, Shuhei Kurita and Daisuke Kawahara	154
<i>Referring Expressions in Human-Robot Common Ground: A Thesis Proposal</i> Jaap Kruijt	161
<i>Source Code is a Graph, Not a Sequence: A Cross-Lingual Perspective on Code Clone Detection</i> Mohammed Ataur Rahman and Julia Ive	168
<i>Distilling Text Style Transfer With Self-Explanation From LLMs</i> Chiyu Zhang, Honglong Cai, Yuezhong Li, Yuexin Wu, Le Hou and Muhammad Abdul-Mageed	200
<i>Reinforcement Learning for Edit-Based Non-Autoregressive Neural Machine Translation</i> Hao Wang, Tetsuro Morimura, Ukyo Honda and Daisuke Kawahara	212
<i>Evaluation Dataset for Japanese Medical Text Simplification</i> Koki Horiguchi, Tomoyuki Kajiwara, Yuki Arase and Takashi Ninomiya	219
<i>Multi-Source Text Classification for Multilingual Sentence Encoder with Machine Translation</i> Reon Kajikawa, Keiichiro Yamada, Tomoyuki Kajiwara and Takashi Ninomiya	226
<i>A Reproducibility Study on Quantifying Language Similarity: The Impact of Missing Values in the URIEL Knowledge Base</i> Hasti Toossi, Guo Qing Huai, Jinyu Liu, Eric Khiu, A. Seza Dođruöz and En-Shiun Annie Lee	233
<i>Coding Open-Ended Responses using Pseudo Response Generation by Large Language Models</i> Yuki Zenimoto, Ryo Hasegawa, Takehito Utsuro, Masaharu Yoshioka and Noriko Kando	242
<i>Cross-Task Generalization Abilities of Large Language Models</i> Qinyuan Ye	255
<i>Commentary Generation from Data Records of Multiplayer Strategy Esports Game</i> Zihan Wang and Naoki Yoshinaga	263
<i>Facilitating Opinion Diversity through Hybrid NLP Approaches</i> Michiel Van Der Meer	272
<i>HybridBERT - Making BERT Pretraining More Efficient Through Hybrid Mixture of Attention Mechanisms</i> Gokul Srinivasagan and Simon Ostermann	285