

9th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL-2025)

Albuquerque, New Mexico, USA
4 May 2025

ISBN: 979-8-3313-1921-2

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>Matching and Linking Entries in Historical Swedish Encyclopedias</i> Simon Börjesson, Erik Ersmark and Pierre Nugues	1
<i>Preserving Comorian Linguistic Heritage: Bidirectional Transliteration Between the Latin Alphabet and the Kamar-Eddine System</i> Abdou Mohamed Naira, Abdessalam Bahafid, Zakarya Erraji, Anass Allak, Mohamed Soibira Naoufal and Imade Benelallam	11
<i>LLM-based Adversarial Dataset Augmentation for Automatic Media Bias Detection</i> Martin Wessel	19
<i>HieroLM: Egyptian Hieroglyph Recovery with Next Word Prediction Language Model</i> Xuheng Cai and Erica Zhang	25
<i>Evaluating LLM-Prompting for Sequence Labeling Tasks in Computational Literary Studies</i> Axel Pichler, Janis Pagel and Nils Reiter	32
<i>Generation of Russian Poetry of Different Genres and Styles Using Neural Networks with Character-Level Tokenization</i> Ilya Koziev and Alena Fenogenova	47
<i>Automating Violence Detection and Categorization from Ancient Texts</i> Alhassan Abdelhalim and Michaela Regneri	64
<i>Rethinking Scene Segmentation. Advancing Automated Detection of Scene Changes in Literary Texts</i> Svenja Guhr, Huijun Mao and Fengyi Lin	79
<i>Sentence-Alignment in Semi-parallel Datasets</i> Steffen Frenzel and Manfred Stede	87
<i>Argumentation in political empowerment on Instagram</i> Aenne Knierim and Ulrich Heid	97
<i>Interpretable Models for Detecting Linguistic Variation in Russian Media: Towards Unveiling Propagandistic Strategies during the Russo-Ukrainian War</i> Anastasiia Vestel and Stefania Degaetano-Ortlieb	109
<i>Tuning Into Bias: A Computational Study of Gender Bias in Song Lyrics</i> Danqing Chen, Adithi Satish, Rasul Khanbayov, Carolin Schuster and Georg Groh	117
<i>Artificial Relationships in Fiction: A Dataset for Advancing NLP in Literary Domains</i> Despina Christou and Grigorios Tsoumakas	130
<i>Improving Hate Speech Classification with Cross-Taxonomy Dataset Integration</i> Jan Fillies and Adrian Paschke	148
<i>Classifying Textual Genre in Historical Magazines (1875-1990)</i> Vera Danilova and Ylva Söderfeldt	160
<i>Lexical Semantic Change Annotation with Large Language Models</i> Thora Hagen	172
<i>AI Conversational Interviewing: Transforming Surveys with LLMs as Adaptive Interviewers</i> Alexander Wuttke, Matthias Assenmacher, Christopher Klamm, Max Lang and Fraue Kreuter	179

<i>Embedded Personalities: Word Embeddings and the Big Five Personality Model</i>	
Oliver Müller and Stefania Degaetano-Ortlieb	205
<i>Prompting the Past: Exploring Zero-Shot Learning for Named Entity Recognition in Historical Texts Using Prompt-Answering LLMs</i>	
Crina Tudor, Beata Megyesi and Robert Östling	216
<i>LLMs for Translation: Historical, Low-Resourced Languages and Contemporary AI Models</i>	
Merve Tekgürler	227
<i>Optimizing Cost-Efficiency with LLM-Generated Training Data for Conversational Semantic Frame Analysis</i>	
Shiho Matta, Yin Jou Huang, Fei Cheng, Hirokazu Kiyomaru and Yugo Murawaki	238
<i>Don't stop pretraining! Efficiently building specialised language models in resource-constrained settings.</i>	
Sven Najem-Meyer, Frédéric Kaplan and Matteo Romanello	252
<i>'... like a needle in a haystack': Annotation and Classification of Comparative Statements</i>	
Pritha Majumdar, Franziska Pannach, Arianna Graciotti and Johan Bos	261
<i>Identifying Small Talk in Natural Conversations</i>	
Steffen Frenzel and Annette Hautli-Janisz	272
<i>Why Novels (Don't) Break Through: Dynamics of Canonicity in the Danish Modern Breakthrough (1870-1900)</i>	
Alie Lassche, Pascale Feldkamp, Yuri Bizzoni, Katrine Baunvig and Kristoffer Nielbo	278
<i>Adapting Multilingual Embedding Models to Historical Luxembourgish</i>	
Andrianos Michail, Corina Raclé, Juri Opitz and Simon Clematide	291