

8th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL-2024)

St. Julian's, Malta
22 March 2024

ISBN: 978-1-7138-9441-4

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

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>Evaluating In-Context Learning for Computational Literary Studies: A Case Study Based on the Automatic Recognition of Knowledge Transfer in German Drama</i> Janis Pagel, Axel Pichler and Nils Reiter	1
<i>Coreference in Long Documents using Hierarchical Entity Merging</i> Talika Gupta, Hans Ole Hatzel and Chris Biemann	11
<i>Metaphorical Framing of Refugees, Asylum Seekers and Immigrants in UKs Left and Right-Wing Media</i> Yunxiao Wang	18
<i>Computational Analysis of Dehumanization of Ukrainians on Russian Social Media</i> Kateryna Burovova and Mariana Romanyshyn	28
<i>Compilation of a Synthetic Judeo-French Corpus</i> Iglika Nikolova-Stoupak, Gaél Lejeune and Eva Schaeffer-Lacroix	40
<i>Detecting Structured Language Alternations in Historical Documents by Combining Language Identification with Fourier Analysis</i> Hale Sirin, Sabrina Li and Thomas Lippincott	46
<i>EmotionArcs: Emotion Arcs for 9,000 Literary Texts</i> Emily Ohman, Yuri Bizzoni, Pascale Feldkamp Moreira and Kristoffer Nielbo	51
<i>Multi-word Expressions in English Scientific Writing</i> Diego Alves, Stefan Fischer, Stefania Degaetano-Ortlieb and Elke Teich	67
<i>EventNet-ITA: Italian Frame Parsing for Events</i> Marco Rovera	77
<i>Modeling Moravian Memoirs: Ternary Sentiment Analysis in a Low Resource Setting</i> Patrick Brookshire and Nils Reiter	91
<i>Applying Information-theoretic Notions to Measure Effects of the Plain English Movement on English Law Reports and Scientific Articles</i> Sergei Bagdasarov and Stefania Degaetano-Ortlieb	101
<i>Uncovering the Handwritten Text in the Margins: End-to-end Handwritten Text Detection and Recognition</i> Liang Cheng, Jonas Frankemölle, Adam Axelsson and Ekta Vats	111
<i>Historical Portrayal of Greek Tourism through Topic Modeling on International Newspapers</i> Eirini Karamouzi, Maria Pontiki and Yannis Krasnikoulakis	121
<i>Post-Correction of Historical Text Transcripts with Large Language Models: An Exploratory Study</i> Emanuela Boros, Maud Ehrmann, Matteo Romanello, Sven Najem-Meyer and Frédéric Kaplan	133
<i>Distinguishing Fictional Voices: a Study of Authorship Verification Models for Quotation Attribution</i> Gaspard Michel, Elena Epure, Romain Hennequin and Christophe Cerisara	160
<i>Perplexing Canon: A study on GPT-based perplexity of canonical and non-canonical literary works</i> Yaru Wu, Yuri Bizzoni, Pascale Moreira and Kristoffer Nielbo	172

<i>People and Places of the Past - Named Entity Recognition in Swedish Labour Movement Documents from Historical Sources</i>	
Crina Tudor and Eva Pettersson	185
<i>Part-of-Speech Tagging of 16th-Century Latin with GPT</i>	
Elina Stüssi and Phillip Ströbel	196
<i>Two Approaches to Diachronic Normalization of Polish Texts</i>	
Kacper Dudzic, Filip Gralinski, Krzysztof Jassem, Marek Kubis and Piotr Wierzchoń	207
<i>Enriching the Metadata of Community-Generated Digital Content through Entity Linking: An Evaluative Comparison of State-of-the-Art Models</i>	
Youssef Benkhedda, Adrians Skapars, Viktor Schlegel, Goran Nenadic and Riza Batista-Navarro	213
<i>Recognising Occupational Titles in German Parliamentary Debates</i>	
Johanna Binnewitt	221
<i>Dynamic embedded topic models and change-point detection for exploring literary-historical hypotheses</i>	
Hale Sirin and Thomas Lippincott	231
<i>Post-OCR Correction of Digitized Swedish Newspapers with ByT5</i>	
Viktoria Löfgren and Dana Dannélls	237
<i>The Kronieken Corpus: an Annotated Collection of Dutch/Flemish Chronicles from 1500-1850</i>	
Theo Dekker, Erika Kuijpers, Alie Lassche, Carolina Lenarduzzi, Roser Morante and Judith Pollmann	243
<i>Direct Speech Identification in Swedish Literature and an Exploration of Training Data Type, Typographical Markers, and Evaluation Granularity</i>	
Sara Stymne	253
<i>Pairing Orthographically Variant Literary Words to Standard Equivalents Using Neural Edit Distance Models</i>	
Craig Messner and Thomas Lippincott	264
<i>[Lions: 1] and [Tigers: 2] and [Bears: 3], Oh My! Literary Coreference Annotation with LLMs</i>	
Rebecca Hicke and David Mimno	270
<i>Stage Direction Classification in French Theater: Transfer Learning Experiments</i>	
Alexia Schneider and Pablo Ruiz Fabo	278