

# **5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature (LaTeCH-CLfL-2021)**

Held online

Punta Cana, Dominican Republic  
11 November 2021

ISBN: 978-1-7138-9175-8

**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© (2021) 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>The Early Modern Dutch Mediascape. Detecting Media Mentions in Chronicles Using Word Embeddings and CRF</i>                              |     |
| Alie Lassche and Roser Morante .....  | 1   |
| <i>FrameNet-like Annotation of Olfactory Information in Texts</i>   |     |
| Sara Tonelli and Stefano Menini .....   | 11  |
| <i>Batavia asked for advice. Pretrained language models for Named Entity Recognition in historical texts.</i>                               |     |
| Sophie I. Arnoult, Lodewijk Petram and Piek Vossen .....  | 21  |
| <i>Quantifying Contextual Aspects of Inter-annotator Agreement in Intertextuality Research</i>  |     |
| Enrique Manjavacas Arevalo, Laurence Mellerin and Mike Kestemont .....  | 31  |
| <i>The Multilingual Corpus of Survey Questionnaires Query Interface</i>   |     |
| Danielly Sorato and Diana Zavala-Rojas .....  | 43  |
| <i>The FairyNet Corpus - Character Networks for German Fairy Tales</i>  |     |
| David Schmidt, Albin Zehe, Janne Lorenzen, Lisa Sergel, Sebastian Düker, Markus Krug and Frank Puppe .....                                  | 49  |
| <i>End-to-end style-conditioned poetry generation: What does it take to learn from examples alone?</i>                                      |     |
| Jörg Wöckener, Thomas Haider, Tristan Miller, The-Khang Nguyen, Thanh Tung Linh Nguyen, Minh Vu Pham, Jonas Belouadi and Steffen Eger ..... | 57  |
| <i>Emotion Classification in German Plays with Transformer-based Language Models Pretrained on Historical and Contemporary Language</i>     |     |
| Thomas Schmidt, Katrin Dennerlein and Christian Wolff .....   | 67  |
| <i>Automating the Detection of Poetic Features: The Limerick as Model Organism</i>  |     |
| Almas Abdibayev, Yohei Igarashi, Allen Riddell and Daniel Rockmore .....  | 80  |
| <i>Unsupervised Adverbial Identification in Modern Chinese Literature</i>   |     |
| Wenxiu Xie, John Lee, Fangqiong Zhan, Xiao Han and Chi-Yin Chow .....   | 91  |
| <i>Data-Driven Detection of General Chiasmi Using Lexical and Semantic Features</i>   |     |
| Felix Schneider, Björn Barz, Phillip Brandes, Sophie Marshall and Joachim Denzler .....   | 96  |
| <i>Translationese in Russian Literary Texts</i>   |     |
| Maria Kunilovskaya, Ekaterina Lapshinova-Koltunski and Ruslan Mitkov .....  | 101 |
| <i>BAHP: Benchmark of Assessing Word Embeddings in Historical Portuguese</i>  |     |
| Zuoyu Tian, Dylan Jarrett, Juan Escalona Torres and Patricia Amaral .....   | 113 |
| <i>The diffusion of scientific terms – tracing individuals’ influence in the history of science for English</i>                             |     |
| Yuri Bizzoni, Stefania Degaetano-Ortlieb, Katrin Menzel and Elke Teich .....  | 120 |
| <i>A Pilot Study for BERT Language Modelling and Morphological Analysis for Ancient and Medieval Greek</i>                                  |     |
| Pranaydeep Singh, Gorik Rutten and Els Lefever .....  | 128 |
| <i>Zero-Shot Information Extraction to Enhance a Knowledge Graph Describing Silk Textiles</i>   |     |
| Thomas Schleider and Raphael Troncy .....   | 138 |

|  |     |
|--|-----|
| <i>'Technologica cosa': Modeling Storyteller Personalities in Boccaccio's 'Decameron'</i>                                      |     |
| A. Cooper, Maria Antoniak, Christopher De Sa, Marilyn Migiel and David Mimno .....   | 147 |
| <i>WMDecompose: A Framework for Leveraging the Interpretable Properties of Word Mover's Distance in Sociocultural Analysis</i> |     |
| Mikael Brunila and Jack LaViolette .....   | 154 |
| <i>Period Classification in Chinese Historical Texts</i>   |     |
| Zuoyu Tian and Sandra Kübler .....   | 168 |
| <i>A Mixed-Methods Analysis of Western and Hong Kong-based Reporting on the 2019–2020 Protests</i>                             |     |
| Arya D. McCarthy, James Scharf and Giovanna Maria Dora Dore .....  | 178 |
| <i>Stylometric Literariness Classification: the Case of Stephen King</i>   |     |
| Andreas van Cranenburgh and Erik Ketzan .....  | 189 |