

EACL 2023

**The 7th Joint SIGHUM Workshop on Computational  
Linguistics for Cultural Heritage, Social Sciences, Humanities  
and Literature**

**Proceedings of LaTeCH-CLfL 2023**

May 5, 2023

©2023 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
209 N. Eighth Street  
Stroudsburg, PA 18360  
USA  
Tel: +1-570-476-8006  
Fax: +1-570-476-0860  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 978-1-959429-54-8

## Introduction

Welcome to our workshop! If you have been here earlier: good to have you back. If this is your first visit: make yourself at home. This event, which has been around in a few forms for nearly two decades, always covers a tantalizingly wide palette of topics. It is not different this year.

First off, there is the usual helping of articles on the broad subject of literature. There are papers on literary devices: antonomasia (you want to find out more), rhymes and adverbial markers. Also, quote detection and quote attribution; authorship verification; a look at models of humour; an analysis of emotional narratives; scent mining (a study of olfactory information in texts).

Next, we have three papers on ancient languages, Latin twice and Greek once; and two papers on the diachronic study of language – English science writing over 300 years, and the gender of book authors over 200 years.

Then there is a paper on biases in Named Entity Recognition, two papers on text and current politics (the European refugee crisis and the Russian invasion of Ukraine), and work on exploring Social Science archives via question-answering.

Last but not least, it turns out not only that action on climate change costs a colossal amount of money but that such money can be the subject of incorrect reporting. We have a paper just about that.

As you can see, there is something for everyone (all things considered) but do keep an open mind and read all nineteen papers. You will be glad you did.

Do not forget to visit our Web site [HERE](#) – and check out past workshops too.

Stefania, Anna, Nils, Stan

# The program

## 9:00-10:30: Talks I

*Named Entity Annotation Projection Applied to Classical Languages*  
Tariq Yousef, Chiara Palladino, Gerhard Heyer and Stefan Jänicke

*Linking the Neulateinische Wortliste to the anonymous Knowledge Base of Interoperable Resources for Latin*  
Federica Iurescia, Eleonora Litta, Marco Passarotti, Matteo Pellegrini, Giovanni Moretti and Paolo Ruffolo

*Fractality of informativity in 300 years of English scientific writing*  
Yuri Bizzoni and Stefania Degaetano-Ortlieb

## 10:30-11:15: Coffee break / SIGHUM Business Meeting

## 11:15-12:45: Talks II

*GPoeT: a Language Model Trained for Rhyme Generation on Synthetic Data*  
Andrei Popescu-Belis, Àlex R. Atrio, Bastien Bernath, Etienne Boisson, Teo Ferrari, Xavier Theimer-Lienhard and Giorgos Vernikos

*Scent Mining: Extracting Olfactory Events, Smell Sources and Qualities*  
Stefano Menini, Teresa Paccosi, Serra Sinem Tekiroğlu and Sara Tonelli

*“Who is the Madonna of Italian-American Literature?”: Extracting and Analyzing Target Entities of Vossian Antonomasia*  
Michel Schwab, Robert Jäschke and Frank Fischer

## 12:45-14:15: Lunch break

## 14:15-15:45: Posters

*Standard and Non-standard Adverbial Markers: a Diachronic Analysis in Modern Chinese Literature*  
John Lee, Fangqiong Zhan, Wenxiu Xie, Xiao Han, Chi-Yin Chow and Kam-Yiu Lam

*Quote Detection: A New Task and Dataset for NLP*  
Selma Tekir, Aybüke Güzel, Samet Tenekeci and Bekir Haman

*Improving Long-Text Authorship Verification via Model Selection and Data Tuning*  
Trang Nguyen, Charlie Dagli, Kenneth Alperin, Courtland Vandam and Elliot Singer

*Direct Speech Quote Attribution for Dutch Literature*  
Andreas van Cranenburgh and Frank van den Berg

*Large Bibliographies as a Source of Data for the Humanities ? NLP in the Analysis of Gender of Book Authors in German Countries and in Poland (1801-2021)*  
Adam Pawłowski and Tomasz Walkowiak

*Using Text Classification with a Bayesian Correction for Estimating Overreporting in the Creditor Reporting System on Climate Adaptation Finance*

Janos Borst, Thomas Wencker and Andreas Niekler

*Detecting intersectionality in NER models: A data-driven approach*

Ida Marie Lassen, Mina Almasi, Kenneth Enevoldsen and Ross Deans Kristensen-McLachlan

*OdyCy – A general-purpose NLP pipeline for Ancient Greek*

Jan Kostkan, Márton Kardos, Jacob Palle Bliddal Mortensen and Kristoffer Laigaard Nielbo

*Exploring Social Sciences Archives with Explainable Document Linkage through Question Generation*

Elie Antoine, Hyun Jung Kang, Ismaël Rousseau, Ghislaine Azémard, Frederic Bechet and Geraldine Damnati

*Towards a More In-Depth Detection of Political Framing*

Qi Yu

*Our kind of people? Detecting populist references in political debates* (paper from Findings)

Christopher Klamm, Ines Rehbein and Simone Paolo Ponzetto

*Bridging Argument Quality and Deliberative Quality Annotations with Adapters* (paper from Findings)

Neele Falk and Gabriella Lapesa

**15:45-16:30: Coffee break**

**16:30-18:00: Talks III**

*Emotion Recognition based on Psychological Components in Guided Narratives for Emotion Regulation*

Gustave Cortal, Alain Finkel, Patrick Paroubek and Lina Ye

*Wartime Media Monitor (WarMM-2022): A Study of Information Manipulation on Russian Social Media during the Russia-Ukraine War*

Maxim Alyukov, Maria Kunilovskaya and Andrei Semenov

*What do Humor Classifiers Learn? An Attempt to Explain Humor Recognition Models*

Marcio Lima Inácio, Gabriela Wick-Pedro and Hugo Goncalo Oliveira

# Organizing Committee

## Organizers

Stefania Degaetano-Ortlieb, Saarland University  
Anna Kazantseva, National Research Council Canada  
Nils Reiter, University of Cologne  
Stan Szpakowicz, EECS, University of Ottawa

## Program Committee

### Program Committee

Beatrice Alex, University of Edinburgh, Edinburgh Futures Institute, School of Literatures, Languages and Cultures, School of Informatics  
Melanie Andresen, Universität Stuttgart  
Jinyeong Bak, Sungkyunkwan University  
Yuri Bizzoni, Aarhus University  
Paul Buitelaar, University of Galway  
Miriam Butt, University of Konstanz  
Thierry Declerck, DFKI GmbH  
Stefanie Dipper, Ruhr-Universität Bochum  
Jacob Eisenstein, Google  
Anna Feldman, Montclair State University  
Mark Finlayson, FIU  
Heather Froehlich, The Pennsylvania State University  
Francesca Frontini, Istituto di Linguistica Computazionale A. Zampolli" - ILC Consiglio Nazionale delle Ricerche - CNR  
Udo Hahn, Friedrich-Schiller-Universität Jena  
Serge Heiden, ENS de Lyon  
Labiba Jahan, Augustana College  
Fotis Jannidis, Universität Würzburg  
Dimitrios Kokkinakis, University of Gothenburg  
Stasinos Konstantopoulos, NCSR Demokritos  
Markus Krug, University of Würzburg  
John Lee, City University of Hong Kong  
Chaya Liebeskind, Jerusalem College of Technology , Lev Academic Center  
Thomas Lippincott, Johns Hopkins University  
Barbara McGillivray, King's College London  
David Mimno, Cornell University  
Syrielle Montariol, EPFL  
Vivi Nastase, University of Geneva  
Borja Navarro-colorado, University of Alicante  
Claes Neufeind, University of Cologne  
Kristoffer Nielbo, Center for Humanities Computing, Aarhus University  
Pierre Nugues, Lund University  
Petya Osenova, Sofia University St. Kl. Ohridski and IICT-BAS  
Janis Pagel, Department of Digital Humanities, University of Cologne  
Andrew Piper, McGill University  
Petr Plechac, Institute of Czech Literature CAS  
Thierry Poibeau, LATTICE (CNRS and ENS/PSL)  
Jelena Prokic, Leiden University  
Georg Rehm, DFKI  
Martin Reynaert, ILLC - Universiteit van Amsterdam / DCA - Tilburg University  
Pablo Ruiz Fabo, LiLPa, Université de Strasbourg  
Marijn Schraagen, Utrecht University  
Pia Sommerauer, Vrije Universiteit Amsterdam  
Elke Teich, Universität des Saarlandes  
Ulrich Tiedau, University College London

Menno Van Zaanen, South African Centre for Digital Language Resources  
Rob Voigt, Northwestern University  
Albin Zehe, University of Wuerzburg  
Heike Zinsmeister, Universität Hamburg

### **Additional Reviewers**

Kaveh Aryan Poo, King's College London  
Kenneth Enevoldsen, Aarhus University  
Andrea Farina, King's College London  
Olumide Ojo, Instituto Politécnico Nacional



## Table of Contents

<i>Standard and Non-standard Adverbial Markers: a Diachronic Analysis in Modern Chinese Literature</i> John Lee, Fangqiong Zhan, Wenxiu Xie, Xiao Han, Chi-yin Chow and Kam-yiu Lam . . . . .	1
<i>GPoeT: a Language Model Trained for Rhyme Generation on Synthetic Data</i> Andrei Popescu-belis, Àlex R. Atrio, Bastien Bernath, Etienne Boisson, Teo Ferrari, Xavier Theimer-lienhard and Giorgos Vernikos . . . . .	10
<i>Quote Detection: A New Task and Dataset for NLP</i> Selma Tekir, Aybüke Güzel, Samet Tenekeci and Bekir Haman . . . . .	21
<i>Improving Long-Text Authorship Verification via Model Selection and Data Tuning</i> Trang Nguyen, Charlie Dagli, Kenneth Alperin, Courtland Vandam and Elliot Singer . . . . .	28
<i>Fractality of informativity in 300 years of English scientific writing</i> Yuri Bizzoni and Stefania Degaetano-ortlieb . . . . .	38
<i>Direct Speech Quote Attribution for Dutch Literature</i> Andreas Van Cranenburgh and Frank Van Den Berg . . . . .	45
<i>Great Bibliographies as a Source of Data for the Humanities – NLP in the Analysis of Gender of Book Authors in German Countries and in Poland (1801-2021)</i> Adam Pawłowski and Tomasz Walkowiak . . . . .	63
<i>Emotion Recognition based on Psychological Components in Guided Narratives for Emotion Regulation</i> Gustave Cortal, Alain Finkel, Patrick Paroubek and Lina Ye . . . . .	72
<i>Linking the Neulateinische Wortliste to the anonymous Knowledge Base of Interoperable Resources for Latin</i> Federica Iurescia, Eleonora Litta, Marco Passarotti, Matteo Pellegrini, Giovanni Moretti and Paolo Ruffolo . . . . .	82
<i>What do Humor Classifiers Learn? An Attempt to Explain Humor Recognition Models</i> Marcio Inácio, Gabriela Wick-pedro and Hugo Goncalo Oliveira . . . . .	88
<i>Constructing a Credible Estimation for Overreporting of Climate Adaptation Funds in the Creditor Reporting System</i> Janos Borst, Thomas Wencker and Andreas Niekler . . . . .	99
<i>Who is the Madonna of Italian-American Literature?": Target Entity Extraction and Analysis of Vossian Antonomasia</i> Michel Schwab, Robert Jäschke and Frank Fischer . . . . .	110
<i>Detecting intersectionality in NER models: A data-driven approach</i> Ida Marie Lassen, Mina Almasi, Kenneth Enevoldsen and Ross Kristensen-mclachlan . . . . .	116
<i>OdyCy – A general-purpose NLP pipeline for Ancient Greek</i> Jan Kostkan, Márton Kardos, Jacob Palle Bliddal Mortensen and Kristoffer Laigaard Nielbo . . . . .	128
<i>Scent Mining: Extracting Olfactory Events, Smell Sources and Qualities</i> Stefano Menini, Teresa Paccosi, Serra Sinem Tekiroğlu and Sara Tonelli . . . . .	135
<i>Exploring Social Sciences Archives with Explainable Document Linkage through Question Generation</i> Elie Antoine, Hyun Jung Kang, Ismaël Rousseau, Ghislaine Azémard, Frederic Bechet and Geraldine Damnati . . . . .	141

<i>Wartime Media Monitor (WarMM-2022): A Study of Information Manipulation on Russian Social Media during the Russia-Ukraine War</i>	
Maxim Alyukov, Maria Kunilovskaya and Andrei Semenov .....	152
<i>Towards a More In-Depth Detection of Political Framing</i>	
Qi Yu .....	162
<i>Named Entity Annotation Projection Applied to Classical Languages</i>	
Tariq Yousef, Chiara Palladino, Gerhard Heyer and Stefan Jänicke .....	175