

COLING

**International Conference on
Computational Linguistics**

Proceedings of the Conference and Workshops

COLING

Volume 29 (2022), No. 3

**The 6th Joint SIGHUM Workshop
on Computational Linguistics for Cultural Heritage,
Social Sciences, Humanities and Literature
(LaTeCH-CLfL 2022)**

**The 29th International Conference on
Computational Linguistics**

October 12-17, 2022
Gyeongju, Republic of Korea

Copyright of each paper stays with the respective authors (or their employers).

ISSN 2951-2093

A word from the organizers

Here we go again. This is our sixth joint workshop; also, sixteenth LaTeCH and eleventh CLfL meeting. We have a fine program for you, with something for everyone in the community. We will not single out any paper: they are all worth a good look. There are repeat visitors (welcome back) and new contributors (hello). And Jacob Eisenstein, who has been on the program committee the past five years, is our invited speaker.

No preface to LaTeCH-CLfL can be complete without a huge thank-you to our wonderful program committee. And thanks to all those who submitted their papers.

Stefania, Anna, Nils, Stan

PS. Our workshop's Web site at

<https://sighum.wordpress.com/events/latech-clf1-2022/>
shows all the details. In particular, there is more about the guest speaker and his talk.

Program committee

Melanie Andresen, Hamburg University, Germany
JinYeong Bak, Sungkyunkwan University, Suwon, South Korea
Yuri Bizzoni, Aarhus University, Denmark
Anne-Sophie Bories, Universität Basel, Switzerland
Paul Buitelaar, National University of Ireland, Galway, Ireland
Miriam Butt, University of Konstanz, Germany
Thierry Declerck, Deutsche Forschungszentrum für Künstliche Intelligenz, Germany
Stefanie Dipper, Ruhr-Universität, Bochum, Germany
Jacob Eisenstein, Georgia Institute of Technology, United States
Anna Feldman, Montclair State University, United States
Mark Finlayson, Florida International University, United States
Antske Fokkens, Vrije Universiteit Amsterdam, The Netherlands
Heather Froehlich, Pennsylvania State University, United States
Francesca Frontini, National Research Council, Italy
Udo Hahn, JULIE Lab, Jena, Germany
Mika Härmäläinen, University of Helsinki, Finland
Serge Heiden, École normale supérieure de Lyon, France
Labiba Jahan, Florida International University, United States
Fotis Jannidis, Würzburg University, Germany
Mike Kestemont, University of Antwerp, Belgium
Dimitrios Kokkinakis, University of Gothenburg, Sweden
Stasinou Konstantopoulos, National Centre of Scientific Research "Demokritos", Greece
Markus Krug, Würzburg University, Germany
John Lee, City University of Hong Kong, Hong Kong
Chaya Liebeskind, Jerusalem College of Technology, Israel
Tom Lippincott, Johns Hopkins University, United States
Barbara McGillivray, The Alan Turing Institute, United Kingdom
David Mimno, Cornell University, United States
Syrielle Montariol, Jozef Stefan Institute, Slovenia
Vivi Nastase, University of Stuttgart, Germany
Borja Navarro Colorado, University of Alicante, Spain
Claes Neufeind, University of Cologne, Germany
Kristoffer Laigaard Nielbo, Aarhus University, Denmark
Pierre Nugues, Lund University, Sweden
Petya Osenova, Sofia University and IICT-BAS, Bulgaria
Janis Pagel, University of Cologne, Germany
Andrew Piper, McGill University, Canada
Petr Plecháč, Institute of Czech Literature of the CAS, Czechia
Thierry Poibeau, CNRS Paris and Lattice, France
Jelena Prokic, Leiden University Centre for Digital Humanities, The Netherlands
Georg Rehm, DFKI, Germany
Martin Reynaert, Tilburg University, Radboud University Nijmegen, The Netherlands
Pablo Ruiz Fabo, Université de Strasbourg, France
Marijn Schraagen, Utrecht University, The Netherlands
Matthew Sims, University of California, Berkeley, United States
Pia Sommerauer, Vrije Universiteit Amsterdam, The Netherlands
Elke Teich, Saarland University, Germany
Laure Thompson, University of Massachusetts Amherst, United States
Ulrich Tiedau, University College London, United Kingdom

Ted Underwood, University of Illinois, Urbana-Champaign, United States

Menno van Zaanen, South African Centre for Digital Language Resources, Potchefstroom, South Africa

Albin Zehe, University of Würzburg, Germany

Heike Zinsmeister, University of Hamburg, Germany

Organizers

Stefania Degaetano-Ortlieb

Department of Language Science and Technology, Saarland University

Anna Kazantseva

National Research Council Canada

Nils Reiter

Department of Digital Humanities, University of Cologne

Stan Szpakowicz

School of Electrical Engineering and Computer Science, University of Ottawa

Table of Contents

<i>Evaluation of Word Embeddings for the Social Sciences</i> Ricardo Schiffers, Dagmar Kern and Daniel Hienert	1
<i>Developing a tool for fair and reproducible use of paid crowdsourcing in the digital humanities</i> Tuomo Hiippala, Helmiina Hotti and Rosa Suviranta	7
<i>Archive TimeLine Summarization (ATLS): Conceptual Framework for Timeline Generation over Historical Document Collections</i> Nicolas Gutehrlé, Antoine Doucet and Adam Jatowt	13
<i>Prabhupadavani: A Code-mixed Speech Translation Data for 25 Languages</i> Jivnesh Sandhan, Ayush Daksh, Om Adideva Paranjay, Laxmidhar Behera and Pawan Goyal ...	24
<i>Using Language Models to Improve Rule-based Linguistic Annotation of Modern Historical Japanese Corpora</i> Jerry Bonnell and Mitsunori Ogihara	30
<i>Every picture tells a story: Image-grounded controllable stylistic story generation</i> Holy Lovenia, Bryan Wilie, Romain Barraud, Samuel Cahyawijaya, Willy Chung and Pascale Fung	40
<i>To the Most Gracious Highness, from Your Humble Servant: Analysing Swedish 18th Century Petitions Using Text Classification</i> Ellinor Lindqvist, Eva Pettersson and Joakim Nivre	53
<i>Automatized Detection and Annotation for Calls to Action in Latin-American Social Media Postings</i> Wassiliki Siskou, Clara Giralt Mirón, Sarah Molina-Raith and Miriam Butt	65
<i>The Distribution of Deontic Modals in Jane Austen’s Mature Novels</i> Lauren Levine	70
<i>Man vs. Machine: Extracting Character Networks from Human and Machine Translations</i> Aleksandra Konovalova and Antonio Toral	75
<i>The COVID That Wasn’t: Counterfactual Journalism Using GPT</i> Sil Hamilton and Andrew Piper	83
<i>War and Pieces: Comparing Perspectives About World War I and II Across Wikipedia Language Communities</i> Ana Smith and Lillian Lee	94
<i>Computational Detection of Narrativity: A Comparison Using Textual Features and Reader Response</i> Max Steg, Karlo H. R. Slot and Federico Pianzola	105
<i>Towards a contextualised spatial-diachronic history of literature: mapping emotional representations of the city and the country in Polish fiction from 1864 to 1939</i> Agnieszka Karlińska, Cezary Rosiński, Jan Wieczorek, Patryk Hubar, Jan Kocoń, Marek Kubis, Stanisław Woźniak, Arkadiusz Margraf and Wiktor Walentynowicz	115
<i>Measuring Presence of Women and Men as Information Sources in News</i> Muitze Zulaika, Xabier Saralegi and Iñaki San Vicente	126

Conference Program

Talks

Regular talks I

Automatized Detection and Annotation for Calls to Action in Latin-American Social Media Postings

Wassiliki Siskou, Clara Giralt Mirón, Sarah Molina-Raith and Miriam Butt

The Distribution of Deontic Modals in Jane Austen's Mature Novels

Lauren Levine

Towards a contextualised spatial-diachronic history of literature: mapping emotional representations of the city and the country in Polish fiction from 1864 to 1939

Agnieszka Karlińska, Cezary Rosiński, Jan Wiczorek, Patryk Hubar, Jan Kocoń, Marek Kubis, Stanisław Woźniak, Arkadiusz Margraf and Wiktor Walentynowicz

Regular talks II

Every picture tells a story: Image-grounded controllable stylistic story generation

Holy Lovenia, Bryan Wilie, Romain Barraud, Samuel Cahyawijaya, Willy Chung and Pascale Fung

Computational Detection of Narrativity: A Comparison Using Textual Features and Reader Response

Max Steg, Karlo H. R. Slot and Federico Pianzola

Measuring Presence of Women and Men as Information Sources in News

Muitze Zulaika, Xabier Saralegi and Iñaki San Vicente

Invited talk

Detecting Intellectual Influence from Dynamic Word Embeddings: Applications to Historical Newspapers and Contemporary Research Articles

Jacob Eisenstein

Posters

Evaluation of Word Embeddings for the Social Sciences

Ricardo Schiffers, Dagmar Kern and Daniel Hienert

Developing a tool for fair and reproducible use of paid crowdsourcing in the digital humanities

Tuomo Hiippala, Helmiina Hotti and Rosa Suviranta

Archive TimeLine Summarization (ATLS): Conceptual Framework for Timeline Generation over Historical Document Collections

Nicolas Gutehrlé, Antoine Doucet and Adam Jatowt

Prabhupadavani: A Code-mixed Speech Translation Data for 25 Languages

Jivnesh Sandhan, Ayush Daksh, Om Adideva Paranjay, Laxmidhar Behera and Pawan Goyal

Using Language Models to Improve Rule-based Linguistic Annotation of Modern Historical Japanese Corpora

Jerry Bonnell and Mitsunori Ogihara

To the Most Gracious Highness, from Your Humble Servant: Analysing Swedish 18th Century Petitions Using Text Classification

Ellinor Lindqvist, Eva Pettersson and Joakim Nivre

Man vs. Machine: Extracting Character Networks from Human and Machine Translations

Aleksandra Konovalova and Antonio Toral

The COVID That Wasn't: Counterfactual Journalism Using GPT

Sil Hamilton and Andrew Piper

War and Pieces: Comparing Perspectives About World War I and II Across Wikipedia Language Communities

Ana Smith and Lillian Lee