

# Proceedings of the Computational Humanities Research Conference 2023: Preface

Fotis Jannidis<sup>1</sup>, Iza Romanowska<sup>2</sup> and Artjoms Šeļa<sup>3</sup>

<sup>1</sup>*Julius-Maximilians-Universität Würzburg, Germany*

<sup>2</sup>*Center for Humanities Computing, Aarhus University, Denmark*

<sup>3</sup>*Institute of Polish Language, Polish Academy of Sciences*

## Computational Humanities Research is Scaling Up and Away

The CHR conference is growing. From its nascent edition three years ago it has gone from a new and fresh but unequivocally small event to a recognised and popular gathering of researchers working with computational methods across different humanities domains. This year more papers were submitted and accepted than ever before, involving more researchers (there may be a correlation at work here!), coming from more countries and - last but not least - representing more disciplines than in the previous years.

With a total of 89 accepted submissions, which includes 49 published papers in the main conference track, Paris will host the largest CHR conference to date.

Scaling up, be it a conference or a computation, has its price. The work of the program committee is getting harder in that tough decisions had to be made when deciding which papers to accept and which to leave out. Despite the acceptance rate of only 68%, this year, we were for the first time confronted with the choice between access and time. If we were to accept the same number of papers as the last conference, the acceptance rate would drop to under 35%. While everyone would have the chance to see all of the presentations, we would have to reject papers, including many good papers, which would have easily made it to the conference in the previous years. We choose the other path: for the first time, the conference will have a second, parallel track, enabling us to enjoy a higher diversity of papers but also forcing us to make decisions about what to miss out on. While we celebrate the rising popularity of the conference and the rapid growth of the whole field, in the long run, the choice between inclusivity and size is something that our community will have to decide on. To meet this increased demand for participation, we also introduced a lightning talk session, shifted submissions around, and slightly cut the time of the presentation (but not the discussion!), all in the hope that by accepting more papers and hearing from more diverse researchers we can all engage in a more

---

*CHR 2023: Computational Humanities Research Conference, December 6–8, 2023, Paris, France*

✉ [fotis.jannidis@uni-wuerzburg.de](mailto:fotis.jannidis@uni-wuerzburg.de) (F. Jannidis); [iromanowska@aias.au.dk](mailto:iromanowska@aias.au.dk) (I. Romanowska); [artjoms.sela@ijp.pan.pl](mailto:artjoms.sela@ijp.pan.pl) (A. Šeļa)

🌐 <https://jannidis.de> (F. Jannidis); <https://pure.au.dk/portal/da/iromanowska@aias.au.dk> (I. Romanowska); <https://artjomsh.github.io/web/> (A. Šeļa)

🆔 0000-0001-6944-6113 (F. Jannidis); 0000-0002-9487-2111 (I. Romanowska)

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

interdisciplinary and inspirational environment.

In 2023 “scaling up” was also a major theme for large language models that broke into the general public’s awareness with ChatGPT reaching 200 million users by the end of August this year. Similarly to previous editions, CHR2023 holds a dedicated session on LLMs, but now these models are also discussed and used everywhere else, blending seamlessly into routines of literary and historical research and authorship attribution methods. The sudden expansion following the recognition of new possibilities to apply AI’s impressive strengths to humanities research problems has only just started, but the same is true for the pitfalls of their usage and we hope that the CHR community will be at the forefront of advances in both of these topics.

There’s also been a welcome growth in the diversity of topics within the Computational Humanities this year, with a stronger presence of music research, archaeology and cultural evolution. Since its first edition, CHR has been open to research in very diverse fields within (and adjacent to) the Humanities, and we are glad that this connection is preserved this year, too. As our cultural evolution colleagues teach us, it is only through diversity that we can grow and innovate in ways that would not be possible if we were to close ourselves off from the wider academic world. We hope that this will be a trigger for more contributions coming from other disciplines such as social sciences, life sciences or even physics in the coming years.

37% of the accepted papers had women as the first author. This gender representation would be considered a success for most scientific events focused on computational methods. In terms of geographic and cultural diversity we still have a long way to go. We know we can do better and will strive to ensure that CHR is known as a place for the most excellent, diverse and original minds in our field.

## **Acknowledgements**

An event like the CHR conference would not be possible without the generous help, tireless work and continuous support of so many people. We would like to especially thank the conference organizers who worked in the background to keep the community going over the years: Folgert Karsdorp and Melvin Wevers were always ready with good advice, and did most of the work with the timely preparation of the proceedings (which, with dozens of papers, was a real challenge this year!). We want to also thank Alie Lassche, a programme chair from the previous CHR edition, who did an unparalleled job of transferring knowledge and workflows to us from last year; thus sparing us vast amounts of time and effort that could then be allocated to other tasks. Many thanks also to the local organizers Florian Cafiero and Marie Puren, who ensured we had somewhere to go (the venue), something to fuel our bodies (the food) and something to inspire excellence in research (the coffee). They also took care of the conference keynote speakers, additional funding, organization of workshops, communication, the website and a million other tiny fires that inevitably need to be put down when organizing a large-scale international gathering. Special thanks go to the members of the program committee who reviewed the 119 submissions in a short time and with a high level of professionalism. In many cases, you saved the day by doing extra, or emergency reviews. We are also very grateful to the three keynote speakers who agreed to join us in Paris to share their expertise and perspectives.

Finally, a conference like the CHR conference would not be possible without you - the researchers attending the event, your work and your willingness to share ideas, insights and

knowledge. Thanks a lot for being part of the CHR community and we hope you are happy with the company you keep! We are confident we can expect a lively and stimulating event. Let Computational Humanities Research grow big and strong!

---

November 2023  
Fotis Jannidis, Iza Romanowska & Artjoms Šeļa

## Program

CHR2023 was a three-day conference held in Paris, France, on December 6 to 8, 2023. Contributions were organized into 11 sessions and a section with lightning talks, as follows:

### SESSION 1: Historical change

- Kristoffer Nielbo, Jan Kostkan, Katrine F. Baunvig, Ekaterina Borisova and Armin W. Geertz. Oscillation between Contemplation and Revelation - Recurrence and Change in the Life History of Teresa of Ávila
- Vojtech Kase, Adéla Sobotková and Petra Heřmánková. Modeling temporal uncertainty in historical datasets
- Wenyi Shang, Song Chen, Yuqi Chen and Jana Diesner. Structural Characteristics in Historical Networks Reveal Changes in Political Culture: An Example From Northern Song China (960–1127 C.E.)
- Thora Hagen and Erik Ketzan. Introducing Traveling Word Pairs in Historical Semantic Change: A Case Study of Privacy Words in 18th and 19th Century English
- Edgar Dubourg, Andrei Mogoutov and Nicolas Baumard. Using a neural network word embedding model to measure cultural innovation

### SESSION 2A: Language

- Caroline Craig, Kartik Goyal, Gregory Crane, Farnoosh Shamsian, and David A. Smith. Testing the Limits of Neural Sentence Alignment Models on Classical Greek and Latin Texts and Translations
- Yulia Clausen. German Question Tags: A Computational Analysis
- Sara Luxmoore, Pedro Ramaciotti Morales and Jonathan Cardoso-Silva. Emoji, language games and political polarisation
- Alie Lassche, Ruben Ros and Joris Veerbeek. (De)constructing Binarism in Journalism: Automatic Antonym Detection in Dutch Newspaper Articles

### SESSION 2B: History

- Wouter Haverals and Mike Kestemont. The Middle Dutch Manuscripts Surviving from the Carthusian Monastery of Herne (14th century): Constructing an Open Dataset of Digital Transcriptions
- Marijn Koolen, Ronald Sluijter, Rik Hoekstra and Joris Oddens. Formulas and decision-making: the case of the States General of the Dutch Republic
- Charles de Dampierre, Valentin Thouzeau and Nicolas Baumard. Using Online Catalogs to Estimate Economic Development in Classical Antiquity
- Stella Verkijk and Piek Vossen. Sunken Ships Shan't Sail: Ontology Design for Reconstructing Events in the Dutch East India Company Archives

### SESSION 3A: Literature and society

- Jean Barré and Thierry Poibeau. Beyond Canonicity. Modeling Canon/Archive Literary Change in French Fiction

- Ying Zhong, Nicolas Baumard and Valentin Thouzeau. The evolution of romantic love in Chinese fiction in the very long run (618 - 2022): A quantitative approach
- Laurine Vianne, Yoann Dupont and Jean Barré. Gender bias in French literature
- Joris J. Van Zundert, Roel Smeets and Andreas Van Cranenburgh. Putting Dutchcoref To the Test: Character Detection and Gender Dynamics in Contemporary Dutch Novels
- Vincent Sarbach-Pulicani. Profiling anonymous authors in the Corsican autonomist press of the inter-war period

### **SESSION 3B: HTR & Collation**

- Luigi Bambaci and Daniel Stoekl Ben Ezra. Enhancing HTR of Historical Texts through Scholarly Editions: A Case Study from an Ancient Collation of the Hebrew Bible
- David Smith, Jacob Murel, Jonathan Parkes Allen and Matthew Thomas Miller. Automatic Collation for Diversifying Corpora: Commonly Copied Texts as Distant Supervision for Handwritten Text Recognition
- Tara Andrews. Algorithms for the manipulation and transformation of text variant graphs
- Sarah Lang, Bernhard Liebl and Manuel Burghardt. Toward a Computational Historiography of Alchemy: Challenges and Obstacles of Object Detection for Historical Illustrations of Mining, Metallurgy and Distillation in 16th–17th Century Print

### **SESSION 4A: Narrative**

- Andrew Piper, Hao Xu and Eric D. Kolaczyk. Modeling Narrative Revelation
- Julian Häußler and Evelyn Gius. Operationalizing and Measuring Conflict in German Novels
- Leonard Konle, Agnes Hilger and Fotis Jannidis. On character perception and plot structure of romance novels
- Pascale Moreira, Yuri Bizzoni, Emily Öhman and Kristoffer Nielbo. Not just Plot(ting): A Comparison of Two Approaches for Understanding Narrative Text Dynamics
- Feng Zhou and Federico Pianzola. Evaluation and alignment of movie events extracted via machine learning

### **SESSION 4B: Libraries and collections**

- John Walsh, Glen Layne-Worthey, Jacob Jett, Boris Capitanu, Peter Organisciak, Ryan Dubnicek and J. Stephen Downie. “The library is open!”: Open data and an open API for the HathiTrust Digital Library
- Mariona Coll Ardanuy, Federico Nanni, Kaspar Beelen and Luke Hare. The Past is a Foreign Place: Improving Toponym Linking for Historical Newspapers
- Thomas Smits, Wouter Haverals, Mike Kestemont, Loren Verreyen and Mona Allaert. Greetings from! Extracting address information from 100,000 historical picture postcards
- Bayrem Kaabachi and Simon Dumas Primbault. A Topological Data Analysis of Navigation Paths within Digital Libraries

### **SESSION 5A: Authorship attribution**

- Sophie Robert-Hayek, Jacques Istas and Frédérique Rey. Unraveling the Synoptic puzzle: stylometric insights into Luke's potential use of Matthew
- Thibault Clérice and Anthony Glaise. Twenty-One Pseudo-Chrysostoms and more: authorship verification in the patristic world
- Rebecca M. M. Hicke and David Mimno. T5 meets Tybalt: author attribution in Early Modern English drama using large language models
- Juan Barrios, Florian Cafiero and Simon Gabay. Detecting Psychological Disorders with Stylometry: the Case of ADHD in Adolescent Autobiographical Narratives

### **SESSION 5B: LLMs**

- Janos Borst, Jannis Klähn and Manuel Burghardt. Death of the Dictionary? – The Rise of Zero-Shot Sentiment Classification
- Giselle Gonzalez Garcia and Christian Weilbach. If the Sources Could Talk: Evaluating Large Language Models for Research Assistance in History
- Ryan Muther, Mathew Barber and David Smith. Querying the Past: Automatic Source Attribution with Language Models
- Ophir Münz-Manor, Pavel Kaganovich and Elishai Ezra-Tsur. Style Transfer of Modern Hebrew Literature Using Text Simplification and Generative Language Modeling

### **SESSION 6: Audio/Video**

- Christof Weiß and Meinard Mueller. Studying Tonal Evolution of Western Choral Music: A Corpus-Based Strategy
- Harin Lee, Romain Hennequin and Manuel Moussallam. Understanding individual and collective diversity of cultural taste through large-scale music listening events
- Nicolas Ruth, Manuel Burghardt and Bernhard Liebl. From Clusters to Graphs – Toward a Scalable Viewing of News Videos
- Alexandra Barancová, Melvin Wevers and Nanne van Noord. Blind Dates: Examining the Expression of Temporality in Historical Photographs
- Martin Ruskov and Sara Sullam. Towards a Phenomenographic Framework for Exploratory Visual Analysis of Bibliographic Data

### **SESSION 7: Literature**

- Jean-Baptiste Camps, Nicolas Baumard, Pierre-Carl Langlais, Olivier Morin, Thibault Clérice and Jade Norindr. Make Love or War? Monitoring the Thematic Evolution of Medieval French Narratives
- Kirill Maslinsky. How Exactly does Literary Content Depend on Genre? A Case Study of Animals in Children's Literature
- Simone Reborá, Marina Lehmann, Anne Heumann, Wei Ding and Gerhard Lauer. Comparing ChatGPT to Human Raters and Sentiment Analysis Tools for German Children's Literature

- Lyra D'Souza and David Mimno. The Chatbot and the Canon: Poetry Memorization in LLMs
- Ida Marie S. Lassen, Pascale Feldkamp Moreira, Yuri Bizzoni, Mads Rosendahl Thomsen and Kristoffer Nielbo. Persistence of gender asymmetries in book reviews within and across genres

## **LIGHTNING TALKS**

- Shiming Shen: From the television corpus to the web corpus using an automatic visual tool: methods of the CROBORA project.
- Christel Annemieke Romein, Sara Veldhoen, Andreas Wagner and J.C. Romain: AI Unleashed: Testing Automatic Metadata against Human Labeling on Handwritten Sources in City-state Bern (1528-1795)
- Rebecca Sutton Koeser and Lara Buchak: Theorizing risk attitudes and rationality using agent based modeling
- Carolina Cucart Mora, Jan-Olaf Reschke, Harry Hall, Kamilla Lomborg, Christine Hertler, Mehdi Saqalli, Matt Grove and Marie-Hélène Moncel: Agent-based models to understand the spatiotemporal patterns of the earliest occupations of Western Europe
- Oleg Sobchuk and Bret Beheim: Dissecting the Trendline: Explaining Historical Change with the Price Equation
- Vojtech Kase and Petr Pavlas: A Computational Approach to the Cultural Evolution of Cognitive Metaphors in Historical Texts (1517-1716)
- Botond Szemes and Mihály Nagy: Information Flow in Dramatic Texts
- Lith Lefranc: Night tactics. Gender and the creation of the modern urban night from a data-driven perspective (Antwerp, 1870-1940)
- Oleg Sobchuk and Artjoms Šeļa: Computational thematics: Comparing algorithms for clustering the genres of literary fiction
- Joris J. Van Zundert, Marijn Koolen, Carsten Schnober, Eva Viviani and Willem Van Hage: Connecting Reading Impact to Characteristics of Novels: Getting from Proxies to Concrete Features?
- Sarah Tew and Melissa Jerome: Recetas: A Bilingual Newspaper Recipe Website
- Johan Malmstedt: Radio entropy: Experiments with large-scale audio data
- Manuel Anglada-Tort, Harin Lee, Marc Schönwiesner, Minsu Park and Nori Jacoby: Studying a global network of cultural influence across 1,423 cities and 53 nations through large-scale music discovery behavior

## People

### Program Chairs

- Fotis Jannidis, Julius-Maximilians-Universität Würzburg
- Iza Romanowska, Aarhus University
- Artjoms Šeļa, Polish Academy of Sciences

### Organizers

- Florian Cafiero, Institut d'Etudes Politiques de Paris / École nationale des chartes (local organizer)
- Marie Puren, EPITA / Centre Jean Mabillon - École nationale des chartes (local organizer)
- Folgert Karsdorp, KNAW Meertens Institute (CHR organizer)
- Melvin Wevers, University of Amsterdam (CHR organizer)
- Alie Lassche, Leiden University

### Best Paper Committee

- Florian Cafiero, École nationale des chartes
- Thora Hagen, Julius-Maximilians-Universität Würzburg
- Folgert Karsdorp, KNAW Meertens Institute
- Iza Romanowska, Aarhus University
- Oleg Sobchuk, Max Planck Institute for Evolutionary Anthropology

### Program Committee

- Alberto Acerbi, University of Bristol
- Tara Andrews, University of Vienna
- David Bamman, University of California, Berkeley
- Andreas Baumann, University of Vienna
- Kaspar Beelen, School of Advanced Study, University of London
- Yuri Bizzoni, Aarhus University
- Tobias Blanke, University of Amsterdam
- Emanuela Boros, EPFL
- Judith Brottrager, TU Darmstadt
- Annelen Brunner, Institut für deutsche Sprache
- Estelle Bunout, University of Luxembourg
- Manuel Burghardt, Computational Humanities, University of Leipzig
- Florian Cafiero, École nationale des chartes
- Jean-Baptiste Camps, École nationales des chartes
- Frédéric Clavert, C2DH / Université du Luxembourg
- Simon Clemenide, Universität Zürich

- Katherine Crowford, The Cyprus Institute
- Antoine Doucet, University of La Rochelle
- Keli Du, University of Trier
- Marten During, Luxembourg Centre for Contemporary and Digital History (C2DH)
- Maciej Eder, Institute of Polish Language, Polish Academy of Sciences
- Tim Evans, Imperial College London
- Stephanie Evert, Friedrich-Alexander-Universität Erlangen-Nürnberg
- Elena Fernandez Fernandez, University of Zurich
- Lauren Fonteyn, Leiden University Centre for Linguistics
- Evelyn Gius, Technical University Darmstadt
- Thora Hagen, Universität Würzburg
- Ahmed Hamdi, Université de La Rochelle, L3i
- Niels Hansen, Aarhus Institute of Advanced Studies, Aarhus University
- Serge Heiden, ENS de Lyon
- Simon Hengchen, iguanodon.ai & Université de Genève
- Berenike Herrmann, Bielefeld University
- Petra Heřmánková, Aarhus University
- Tobias Hodel, Digital Humanities/Walter Benjamin Kolleg/Universität Bern
- Yuerong Hu, School of Information Sciences, University of Illinois at Urbana-Champaign
- Lars Johnsen, National Library of Norway
- Andres Karjus, The University of Edinburgh
- Folgert Karsdorp, Meertens Institute
- Vojtech Kase, Aarhus University
- Mike Kestemont, CLiPS
- Kristian Kjellmann, Aalborg University
- Leonard Konle, Würzburg University
- Marijn Koolen, Royal Netherlands Academy of Arts and Sciences - Humanities Cluster
- Ross Kristensen-McLachlan, Centre for Humanities Computing, Aarhus University
- Alie Lassche, Leiden University
- Cynthia C. S. Liem, Delft University of Technology
- Laura Mandell, Texas A&M University
- Enrique Manjavacas, Leiden University
- Barbara McGillivray, King's College London
- George Mikros, Department of Italian Language and Literature, National and Kapodistrian University of Athens
- David Mimno, Cornell University
- Eetu Mäkelä, University of Helsinki
- Ben Nagy, IJP PAN
- Clemens Neudecker, Berlin State Library
- Andreas Niekler, Universität Leipzig

- Kristoffer Nielbo, Aarhus University, Center for Humanities Computing Aarhus
- Damien Nouvel, INaLCO
- Pedro Ortiz Suarez, DFKI GmbH
- Marco Passarotti, Catholic University of the Sacred Heart
- Francesco Periti, Università degli Studi di Milano Statale
- Federico Pianzola, University of Groningen
- Michael Piotrowski, University of Lausanne
- Andrew Piper, McGill University
- Thierry Poibeau, LaTTiCe-CNRS
- Julien Randon-Furling, SAMM, Univ. Paris Panthéon-Sorbonne
- Simone Reborá, University of Mainz
- Christian Reul, University of Würzburg
- Allen Riddell, Indiana University Bloomington
- Ramona Roller, ETH Zurich
- Matteo Romanello, Ecole Polytechnique Fédérale de Lausanne
- Laurent Romary, INRIA & HUB-ISDL
- Yann Ryan, University of Helsinki
- Christof Schoech, University of Würzburg
- Marijn Schraagen, Utrecht University
- Stefan Schweter, Bayerische Staatsbibliothek München, Digital Library/Munich Digitization Center
- Roel Smeets, Radboud University
- Oleg Sobchuk, Max Planck Institute for Evolutionary Anthropology (Leipzig)
- Adela Sobotkova, Aarhus University
- Elena Spadini, University of Basel
- Rachele Sprugnoli, Università degli Studi di Parma
- Zachary Stine, University of Central Arkansas
- Nina Tahmasebi, University of Gothenburg
- Timothy Tangherlini, University of California, Berkeley
- Laure Thompson, University of Massachusetts Amherst
- Lauren Tilton, University of Richmond
- Mikko Tolonen, University of Helsinki
- Ted Underwood, University of Illinois at Urbana-Champaign
- Matteo Valleriani, Max Planck Institute for the History of Science
- Joris Van Zundert, Huygens ING - Royal Netherlands Academy of Arts and Sciences (KNAW)
- Christof Weiss, University of Würzburg
- Melvin Wevers, University of Amsterdam
- Andreas van Cranenburgh, University of Groningen
- Karina van Dalen-Oskam, Huygens ING
- Marieke van Erp, KNAW Humanities Cluster
- Nanne van Noord, University of Amsterdam