

Proceedings of the Computational Humanities Research Conference 2021: Preface

Maud Ehrmann¹, Folgert Karsdorp², Melvin Wevers³, Tara Andrews⁴,
Manuel Burghardt⁵, Mike Kestemont⁶, Enrique Manjavacas⁷, Michael Piotrowski⁸
and Joris van Zundert⁹

¹*École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland*

²*KNAW Meertens Institute, Amsterdam, the Netherlands*

³*University of Amsterdam, Amsterdam, the Netherlands*

⁴*University of Vienna, Vienna, Austria*

⁵*University of Leipzig, Leipzig, Germany*

⁶*University of Antwerp, Antwerp, Belgium*

⁷*Leiden University, Leiden, the Netherlands*

⁸*University of Lausanne, Lausanne, Switzerland*

⁹*KNAW Huygens Institute, Amsterdam, the Netherlands*

Two is (Almost) Tradition

It seems very likely that future historians will have to portray 2020 CE as an *annus horribilis*, not least because of the worldwide COVID-19 pandemic. Although this might seem like a small consolation in the grand scheme of things, the same future historians will nevertheless have to acknowledge that 2020 was also the year which saw the first edition of the Computational Humanities Research Workshop, which was a real success.¹ The original plan had been to host the workshop in Amsterdam in the autumn of 2020, bringing physically under one roof the community of Humanities scholars with an orientation towards the computational. The global lockdown that was necessitated by the pandemic quickly wiped those happy prospects off the table. Eventually, and only thanks to the agility of the local organizers, the workshop took place as a virtual event online (18–20 November 2020). In spite of these chaotic circumstances, the organizers, presenters and audience can look back with satisfaction on a smooth series of presentations, lively discussions afterwards, and even a fully-fledged social event during the otherwise lonely evenings of the workshop.

One year later, academia is only tentatively picking up its habits again from the pre-pandemic era and global vaccination programs are still underway. For those reasons, including the uncertainties regarding international travel restrictions, it seemed wiser to postpone a physical – or even hybrid – edition for one more year. Although the resilience of our community is


CHR 2021: Computational Humanities Research Conference, November 17–19, 2021, Amsterdam, The Netherlands

✉ maud.ehrmann@epfl.ch (M. Ehrmann)

🆔 0000-0001-9900-2193 (M. Ehrmann)

© 2021 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)

¹<https://2020.computational-humanities-research.org/>

therefore challenged once more, the 2021 edition of the Computational Humanities Research Conference will be able to boast no less attractive a program than the previous installment.² We were very pleased with both the quantity and quality of the submissions. Each of the 51 submitted papers was carefully peer-reviewed by a minimum of three researchers drawing on the expertise of 74 Program Committee members. We were pleased to be able to accept 29 papers (for an acceptance rate of 56.8%): these were nicely balanced across 14 short and 15 long papers. The fact that many more high-quality papers were submitted than could ultimately be accommodated attests to the keen scholarly interest in and enthusiasm for the CHR conference. The program committee took active steps to optimize the modalities of the submission reviewing procedure and will continue to improve this process in the future.

The topical scope of the published papers is as broad as can be expected from a Humanities-wide event, ranging from the field's traditional strongholds of literary, historical and linguistic research to exciting explorations of multimodality and cryptography. The selection of the three keynote speakers – Iza Romanowska (Aarhus Institute of Advanced Studies), Jacob Eisenstein (Google AI), and Sara Tonelli (Fondazione Bruno Kessler) – reflects the same methodological curiosity and eagerness to engage with neighboring disciplines, such as archaeology or natural language processing. The program included below gives a sense of the pleasing diversity in the field.

Acknowledgements

This second CHR edition would not have been possible without the work and commitment of many. We wish to address very special thanks to Folgert Karsdorp, Barbara McGillivray, Adina Nergheș, and Melvin Wevers for initiating the first CHR workshop in 2020. Next, we are immensely grateful to the members of the program committee for their valuable time and efforts in reviewing the submitted papers and helping build the program. We also thank the maintainers of the EasyChair conference system for providing a platform that handles the submission and review process, as well as the CEUR-WS.org team for offering a free open access publication service. Finally, we thank all authors who submitted to the conference, our three keynote speakers, as well as all attendees of the online event for what will undoubtedly be an inspiring intellectual exchange, to which we all are very much looking forward. Thanks and welcome to all!

October 2021

Maud Ehrmann, Folgert Karsdorp, Melvin Wevers, Tara Andrews, Manuel Burghardt, Mike Kestemont, Enrique Manjavacas, Michael Piotrowski, Joris van Zundert.

²<https://2021.computational-humanities-research.org/>

Program and contributions

CHR2021 was a three-day conference held on November 17 to 19, 2020. Contributions were organized into 9 sessions, as follows:

Wednesday 17 November 2021

Session 1 - Big(gish) Data

- *Entity Matching in Digital Humanities Knowledge Graphs*
Jurian Baas, Mehdi Dastani and Ad Feelders
- *Type- and Token-based Word Embeddings in the Digital Humanities*
Anton Ehrmanntraut, Thora Hagen, Leonard Konle and Fotis Jannidis
- *Probabilistic Analysis of Early Modern British Book Prices*
Iiro Tiihonen, Mikko Tolonen and Leo Lahti

Session 2 - In the Media

- *Navigating the Mise-en-Page: Interpretive Machine Learning Approaches to the Visual Layouts of Multi-Ethnic Periodicals*
Benjamin Lee, Joshua Ortiz Baco, Sarah Salter and Jim Casey
- *Event Flow - How Events shaped the Flow of the News, 1950-1995*
Melvin Wevers, Jan Kostkan and Kristoffer Nielbo
- *Mapping AI issues in media through NLP*
Maxime Crepel, Salomé Do, Jean-Philippe Cointet, Dominique Cardon and Yannis Bouachera

Session 3 - Premodern Texts

- *Obtaining More Expressive Corpus Distributions for Standardized Ancient Languages*
Oliver Hellwig, Sven Sellmer and Sebastian Nehrdich
- *More Data and New Tools. Advances in Parsing the Index Thomisticus Treebank*
Federica Gamba, Marco Passarotti and Paolo Ruffolo
- *Classifying Latin Inscriptions of the Roman Empire: A machine-learning approach*
Vojtech Kase, Petra Heřmánková and Adéla Sobotková
- *Geolocation and Named Entity Recognition in ancient texts: a case study about Ghewond's Armenian History*
Marcella Tambuscio and Tara Lee Andrews

Thursday 18 November 2021

Session 4 - Sound and image

- *Towards Multimodal Computational Humanities. Using CLIP to analyze late-nineteenth century magic lantern slides*
Thomas Smits and Mike Kestemont
- Fabian Moss, Maik Köster, Melinda Femminis, Coline Métrailler and François Bavaud
Digitizing a 19th-century music theory debate for computational analysis
- *Evaluating Heuristics for Audio-Visual Translation*
Timo Baumann and Ashutosh Saboo

Session 5 - Language and literature

- *Zeta & Eta: An exploration and evaluation of two dispersion-based measures of distinctiveness*
Keli Du, Julia Dudar, Cora Rok and Christof Schöch
- *Predicting canonization: Comparing canonization scores based on text-extrinsic and -intrinsic features*
Judith Brottrager, Annina Stahl and Arda Arslan
- *DaCy: A Unified Framework for Danish NLP*
Kenneth Enevoldsen, Lasse Hansen and Kristoffer Nielbo
- *Predicting Structural Elements in German Drama*
Janis Pagel, Nidhi Sihag and Nils Reiter

Session 6 - (Geo)graphical analysis

- *Generic semantic segmentation of historical maps*
Rémi Petitpierre, Frédéric Kaplan and Isabella di Lenardo
- *Station to Station: Linking and enriching historical British railway data*
Mariona Coll Ardanuy, Kaspar Beelen, Jon Lawrence, Katherine McDonough, Federico Nanni, Joshua Rhodes, Giorgia Tolfo and Daniel C.S. Wilson
- Ming Jiang, Yuerong Hu, Glen Worthey, Ryan C Dubniecek, Ted Underwood and J Stephen Downie
Impact of OCR Quality on BERT Embeddings in the Domain Classification of Book Excerpts

Friday 19 November 2021

Session 7 - Text analysis

- *Adjusting scope: a computational approach to case-driven research on semantic change*
Lauren Fonteyn and Enrique Manjavacas Arévalo ;
- *The Challenge of Vernacular and Classical Chinese Cross-Register Authorship Attribution*
Haining Wang, Xin Xie and Allen Riddell
- *The Sentiment of Crypto Art*
Massimo Franceschet

Session 8 - Narrative and Style

- *Detecting Narrativity Across Long Time Scales*
Andrew Piper, Sunyam Bagga, Laura Monteiro, Andrew Yang, Marie Labrosse and Yu Lu Liu
- *Automated Event Annotation in Literary Texts*
Michael Vauth, Hans Ole Hatzel, Evelyn Gius and Chris Biemann
- *Assessing Michael Maier's contributions to Francis Anthony's Apologia (1616) using stylometry*
Sarah Lang

Session 9 - Theatre

- *Beyond Idiolectometry? On Racine's Stylometric Signature*
Simon Gabay
- *'Psyché' as a Rosetta stone? Assessing collaborative authorship in the French 17th century theatre*
Florian Cafiero and Jean-Baptiste Camps
- *Exploring Multimodality of Sentiment Analysis in Plays: A Case Study for a Theater Recording of Emilia Galotti*
Thomas Schmidt and Christian Wolff

People

Organisation committee

- Tara Andrews, *University of Vienna* (conference program)
- Manuel Burghardt, *University of Leipzig* (conference program)
- Giovanni Colavizza, *University of Amsterdam* (organisation)
- Maud Ehrmann, *EPFL* (general organisation, conference program)
- Folgert Karsdorp, *KNAW Meertens Institute* (general organisation, conference program, proceedings)
- Mike Kestemont, *University of Antwerp* (organisation)
- Enrique Manjavacas, *Leiden University* (communication)
- Barbara McGillivray, *King's College London and The Alan Turing Institute* (organisation)
- George Mikros, *University of Athens* (organisation)
- Adina Nerghes, *Wageningen University & Research* (organisation)
- Kristoffer Nielbo, *University of Aarhus* (organisation)
- Michael Piotrowski, *University of Lausanne* (conference program)
- Melvin Wevers, *University of Amsterdam* (general organisation, conference program, communication)
- Joris van Zundert, *KNAW Huygens Institute* (conference program)

Program committee

- Alberto Acerbi, *University of Bristol*
- Maristella Agosti, *University of Padua*
- Tara Andrews, *University of Vienna*
- Adrien Barbaresi, *BBAW*
- Andreas Baumann, *University of Vienna*
- Kaspar Beelen, *The Alan Turing Institute*
- Elli Bleeker, *KNAW Huygens Institute*
- Federica Bologna, *Università di Bologna*
- Emanuela Boros, *Université de La Rochelle*
- Estelle Bunout, *University of Luxembourg*
- Manuel Burghardt, *Computational Humanities, University of Leipzig*
- Frédéric Clavert, *C2DH / Université du Luxembourg*
- Giovanni Colavizza, *University of Amsterdam*
- Mariona Coll Ardanuy, *The Alan Turing Institute*
- Andreas van Cranenburgh, *University of Groningen*

- Antoine Doucet, *University of La Rochelle*
- Marten Düring, *C2DH / Université du Luxembourg*
- Maud Ehrmann, *École Polytechnique Fédérale de Lausanne*
- Marieke van Erp, *KNAW Humanities Cluster*
- Pieter Francois, *University of Oxford*
- Heather Froehlich, *The Pennsylvania State University*
- Simon Hengchen, *University of Gothenburg*
- Leah Henrickson, *University of Leeds*
- Tobias Hodel, *Digital Humanities/Walter Benjamin Kolleg/Universität Bern*
- Laura Hollink, *Vrije Universiteit Amsterdam*
- Fotis Jannidis, *Julius-Maximilians-Universität Würzburg*
- Béatrice Joyeux-Prunel, *Université de Genève*
- Folgert Karsdorp, *KNAW Meertens Institute*
- Mike Kestemont, *University of Antwerp*
- Marijn Koolen, *KNAW Huygens Institute*
- Ross Kristensen-McLachlan, *Centre for Humanities Computing, Aarhus University*
- Cynthia Liem, *Delft University of Technology*
- Cerstin Mahlow, *ZHAW Zurich University of Applied Sciences*
- Enrique Manjavacas, *Leiden University*
- Barbara Mcgillivray, *King's College London and The Alan Turing Institute*
- George Mikros, *University of Athens*
- Jean-Philippe Moreux, *Bibliothèque nationale de France*
- Fabian Moss, *École Polytechnique Fédérale de Lausanne*
- Coline Métrailler, *University of Lausanne*
- Federico Nanni, *The Alan Turing Institute*
- Kristoffer Nielbo, *Aarhus University, Center for Humanities Computing Aarhus*
- Damien Nouvel, *INaLCO*
- Pedro Javier Ortiz Suárez, *Inria / Sorbonne Université*
- Marco Passarotti, *Catholic University of the Sacred Heart*
- Michael Piotrowski, *University of Lausanne*
- Andrew Piper, *McGill University*
- Thierry Poibeau, *LaTTiCe-CNRS*
- Nils Reiter, *Institute of Natural Language Processing, Stuttgart University*
- Allen Riddell, *Indiana University Bloomington*
- Ramona Roller, *ETH Zurich*

- Matteo Romanello, *University of Lausanne*
- Iza Romanowska, *Barcelona Supercomputing Center*
- Laurent Romary, *INRIA & HUB-ISDL*
- Yann Ryan, *Queen Mary, University of London*
- Christof Schoech, *University of Würzburg*
- Stefen Schweter, *Bayerische Staatsbibliothek München, Munich Digital Library*
- Roel Smeets, *Radboud University*
- Oleg Sobchuk, *Max Planck Institute for the Science of Human History*
- Elena Spadini, *University of Lausanne*
- Rachele Sprugnoli, *Università Cattolica del Sacro Cuore*
- Lauren Tilton, *University of Richmond*
- Sara Tonelli, *Fondazione Bruno Kessler*
- Arianna Traviglia, *Italian Institute of Technology*
- Inna Uglanova, *Technical University of Darmstadt*
- Ted Underwood, *University of Illinois at Urbana-Champaign*
- Melvin Wevers, *University of Amsterdam*
- Bruna Wundervald, *Maynooth University Hamilton Institute*
- Eva Zangerle, *University of Innsbruck*
- Joris van Zundert, *KNAW Huygens Institute*