

# Preface

Folgert Karsdorp<sup>a</sup>, Barbara McGillivray<sup>b,c</sup>, Adina Nergheš<sup>d</sup> and Melvin Wevers<sup>e</sup>

<sup>a</sup>Royal Dutch Academy of Arts and Science, Meertens Institute, Amsterdam, the Netherlands

<sup>b</sup>The Alan Turing Institute, London, UK

<sup>c</sup>University of Cambridge, Cambridge, UK

<sup>d</sup>Wageningen University, Wageningen, the Netherlands

<sup>e</sup>Department of History, University of Amsterdam, Amsterdam, the Netherlands

## CHR: A New Venue for Computational Humanities Research

In the humanities and related social sciences, the use of computational, statistical, and mathematical approaches has increased in recent years. The research can be characterized by (i) relying on quantifiable rather than strictly qualitative evidence, (ii) making explicit and formalizing theory with the help of mathematical and computational models, (iii) making statistical inferences about model parameters, (iv) the use of computational and automatic procedures for processing and analyzing data, and (v) searching for computational analogies through which theories from different disciplines can be unified.

And yet, despite the undeniable growth of this new research area, many researchers still feel that there is no suitable research-oriented venue to present and publish their computational work that does not lose sight of questions relevant to the humanities. As such, we aim to fill this niche by building a community for Computational Humanities Research. The Computational Humanities Research (CHR2020) workshop, in which authors present peer-reviewed long and short papers, is part of this community-building effort.

The proceedings contain all papers presented at this first edition of the Computational Humanities Research Workshop. The workshop was originally intended to take place in Amsterdam, but due to the COVID-19 pandemic, we decided to hold the workshop in a virtual format. The online platform selected by the organizers<sup>1</sup> is not only a workshop venue but also an asynchronous platform where this growing international and interdisciplinary community can discuss all ideas and questions related to computational humanities research. The workshop included two keynotes. Dong Nguyen, Assistant Professor at the University of Utrecht gave the first keynote. The second keynote was given by Anne Kandler, Senior Scientist at the Max Planck Institute for Evolutionary Anthropology Leipzig, Germany. The workshop was co-organized by the Humanities Cluster and DHLab of the Royal Dutch Academy of Arts and Science and The Alan Turing Institute.

---


*CHR 2020: Workshop on Computational Humanities Research, November 18–20, 2020, Amsterdam, The Netherlands*

✉ [folgert.karsdorp@meertens.knaw.nl](mailto:folgert.karsdorp@meertens.knaw.nl) (F. Karsdorp)

📞 0000-0002-5958-0551 (F. Karsdorp)

© 2020 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)

<sup>1</sup>See <https://discourse.computational-humanities-research.org>

## Program

CHR2020 was a three-day workshop held on November 18 to 20, 2020. The program committee and organizers accepted 29 of the 49 papers submitted. These papers apply computational methods to topics drawing from a wide range of humanities disciplines. We divided the papers into eight categories:

**Cultural Evolution** Four papers discuss topics related to the field of *Cultural Evolution*.

Topics covered are the cultural accumulation of fan fiction (paper 1), the relationship between form and meaning in poetry and its change over time (paper 2), the relationship between popularity and emotional valence (paper 3), and the estimation of the loss of medieval literature using methods from ecodiversity (paper 4).

**History & Historical Texts** The second category of papers revolves around *History* and the application of computational methods to *Historical texts*. Here we find papers proposing new pipelines for the processing of historical newspaper corpora (papers 26 and 24), research into new formal representations of historical texts (paper 25), a paper on the use of network theory to account for missing data in historical archives (paper 28), the development of a spatio-temporal and semantic information system for medieval sources (paper 29), and an application of distant ‘reading’ to study the iconicity of historical images (paper 27).

**Musicology** In a third category, *Musicology*, we also find papers that discuss historical topics, such as paper 15, in which pitch-class distributions in Anton Webern’s work are discussed. The topics ‘emotion’ and ‘sentiment’ recur in various contributions across categories and disciplines, and also in paper 16, in which a large-scale database for musical sentiment is being developed.

**Linguistics** The fourth group contains contributions that discuss topics related to *Linguistics* and specifically how computational methods can provide new insights in this domain. Paper 20 discusses a linguistic approach to misinformation in Chinese, focusing on the role of emotive language and hyperbole to obtain a reader’s attention. Paper 17 explores the possibility of improving syntactic parsers for low-resource languages by training them on related language varieties. Finally, two papers discuss the application of neural language models to historical texts. Paper 19 employs neural language models to assess their usefulness for addressing questions about functional-semantic change in grammatical categories. And paper 18 investigates how models trained on modern text collections can be adapted for historical texts using transfer learning.

**Literary Studies** Language models, either neural or probabilistic, are also employed in several contributions that discuss topics related to *Literary Studies*. First, paper 8 applies recurrent neural networks to the task of detecting literary intertextuality and text-reuse. Second, Paper 6 approaches the problem of intertextuality from

a different angle, using probabilistic language models in combination with multi-level statistical models to identify factors that advance our understanding of text-reuse. Third, probabilistic mixture models are employed in paper 9 as a means to characterize structural relationships between different discourses, focusing on two religious discourses. The applicability of probabilistic language models to literary texts is further explored in the fourth paper in this category (paper 5). The subject of genre is implicit in many of the contributions that discuss literary topics but it is explicitly addressed in paper 7. The authors examine the organizational structure of fiction in terms of genre and how this structure has changed over time. Finally, the impact of book reviews on reading experience is investigated in paper 10. Here, too, genre is examined explicitly, when the authors investigate the relationship between review impact and book genre.

**Methods** In the *Methods* category, we find three contributions with a strong methodological focus. First, paper 22 presents a new approach to trend estimation that relies on advances in information theory and dynamical systems and applies this approach to social media data. Second, paper 23 develops a new method to measure temporal changes in the social construction of time. Their findings support ideas by conceptual historian Reinhart Koselleck on technological advancements and the experience of time. Paper 21, finally, is a plea for the application of the emergent paradigm of ‘probabilistic programming’ in the humanities. The authors demonstrate how probabilistic programming can assist in the detection, quantification, and description of bias in historical data sets.

**Theory** The final category, *Theory*, contains several papers that deal with theoretical aspects of computational humanities research. First, paper 11 reminds us of the fact that computational research in the humanities is complementary to qualitative approaches, and that an artificial dichotomy between quantitative and qualitative approaches should be avoided. The authors describe a theoretical branch within computational humanities research, which also includes paper 12. This paper discusses a classic topic in computational humanities research: how stories and narratives can be computationally represented. The final two papers focus on generative aspects of recent advances in research about neural networks. Paper 13 addresses the question of what kind of information is learned by neural language models, and how this knowledge is related to the real world. Paper 14 advocates the application of generative machine learning in the humanities, for example, in digital art history, where such models can facilitate the exploration of hidden semantic structures.

## Program Committee

- Barbara McGillivray (Chair) – *University of Cambridge & The Alan Turing Institute, UK*
- Alberto Acerbi – *Brunel University, London, UK*
- Tara Andrews – *University of Vienna, Austria*
- Manuel Burghardt – *Universität Leipzig, Germany*

- Giovanni Colavizza – *University of Amsterdam, the Netherlands*
- Maud Ehrmann – *EPFL, Switzerland*
- Marieke van Erp – *KNAW Humanities Cluster, the Netherlands*
- Lauren Fonteyn – *Leiden University, the Netherlands*
- Pieter Francois – *University of Oxford, UK*
- Leah Henrickson – *Loughborough University, UK*
- Mike Kestemont – *University of Antwerp, Belgium*
- Lauren Klein – *Emory University, USA*
- Cynthia Liem – *Delft University of Technology, the Netherlands*
- Enrique Manjavacas – *University of Antwerp, Belgium*
- George Mikros – *National and Kapodistrian University of Athens, Greece*
- Kristoffer Nielbo – *Aarhus University, Denmark*
- Michael Piotrowski – *University of Lausanne, Switzerland*
- Allen Riddell – *Indiana University Bloomington, USA*
- Iza Romanowska – *Barcelona Supercomputing Center, Spain*
- Tatjana Scheffler – *University of Potsdam, Germany*
- Oleg Sobchuk – *Max Planck Institute for the Science of Human History, Germany*
- Ted Underwood – *University of Illinois, USA*
- Eva Zangerle – *University of Innsbruck, Austria*
- Joris van Zundert – *Huygens ING, the Netherlands*

## Acknowledgements

The organizers of CHR2020 wish to thank the program committee members for their time and effort in reviewing all submitted papers. We would also like to express our gratitude to our keynote speakers Dong Nguyen and Anne Kandler for their engaging contributions. Last but not least, we would like to thank Anneca York, from The Alan Turing Institute, for her technical expertise and support in setting up the live sessions of the workshop.

---

October 2020

CHR2020 Organizers  
Folger Karsdorp, Barbara McGillivray,  
Adina Nerghes, and Melvin Wevers