



Beyond Digital Humanities: How computational methods are reshaping scholarly research

**International Workshop
4 September 2023**

Program

Christian Henriot & Cécile Armand, Aix-Marseille University
Beyond the Digital Humanities: A Position paper

Pierre Magistry, INALCO
Beyond the optimism around huge foundation models in machine learning

Shih-Pei Chen, Max Planck Institute for the History of Science
Fenye by the Numbers: A Quantitative Analysis of Astrological Contents in Chinese Local Gazetteers

Baptiste Blouin, Aix-Marseille University
Bridging NLP and DH: Beyond metrics to meaning

Lena Henningsen, Freiburg University
Reading Practices in the People's Republic of China: Digital Tools developed by the ReadChina Project

Christian Henriot, Aix-Marseille University
Eminent Chinese of the Shenbao: How to 'read' 50,525 newspaper articles

Jiang Jie, Shanghai Normal University
Origin, Genres, and Achievements: The Development of Digital History in Mainland China (2000-2023)



Beyond Digital Humanities: How computational methods are reshaping scholarly research

**International Workshop
4 September 2023**

In the last decade the Digital Humanities (DH) movement has swept the academic landscape in the United States, offering innovative intersections between digital tools and traditional humanities. More recently, in Europe and China, DH has become a new mantra. However, we argue that the real transformative power transcends the broad DH label, rooted in the depth and specificity of computational methodologies. By critically examining examples drawn from disciplines like history, literature, and sociology, we highlight how computational methods offer both macroscopic and microscopic insights, reshaping the very essence of research. The future beckons not a supersession of traditional methods, but a harmonious integration, championing methodological rigor and critical digital literacy.

Presentations

1. Beyond the optimism around huge foundation models in machine learning

Pierre Magistry, INALCO

Abstract

Recent advances in deep learning applied to Natural Language Processing (NLP) have brought significant progress on many tasks, and impressive online demonstration systems and tools are now available to a large audience. Although these models require a very costly pretraining step, they can be shared and adapted to many different NLP tasks. But this progress has partly reshaped our methods to process language data (both spoken and written) in a way that also raise a number of questions and new issues, including normalization, computing power, availability of insanely large corpora, and methods to detect and correct all kinds of unwanted biases. These issues were all important motivations to design our DLS-HN project, a 42-month ANR-funded project which will start by the end of the year.

In this talk, I will present these motivations as an illustration of the issues we are facing in the current state of NLP, especially when applied to DH. DLS-HN aims at exploring these issues under the light of the diversity of Sinitic languages. We will design experiment to study the behaviour of foundation models on our datasets and to assess whether we can propose smaller and more focused models trained on more controlled data. We will challenge the foundation models by working on languages with different levels of proximity with modern standard Mandarin, by either working on more ancient documents or related minority languages (namely Taiwanese and Teochew, two Southern Min languages).

Short bio

Pierre Magistry, an Associate Professor at INALCO, was a key contributor to the ENP-China project, specializing in Natural Language Processing (NLP). Serving from 2018 to 2020, Pierre was instrumental in shaping the project's infrastructure and crafting NLP tools for historians. With a PhD from ALPAGE (Paris Diderot – INRIA) focused on Chinese Word Segmentation, he has extensive experience in Sinitic and low-resource languages, bolstered by multiple research stints in Taiwan.



Pierre's expertise extends to the development of the first mobile input method for the Taiwanese language.

2. Reading Practices in the People's Republic of China: Digital Tools developed by the ReadChina Project

Lena Henningsen, Freiburg University

Abstract

As part of our research into the politics and practices of reading in the People's Republic of China, the ReadChina project has developed a number of computational tools with the aim of archiving data, providing it to the research community and analyzing it in our own case studies. They have proven useful both in quantitative and qualitative analyses. In this presentation, I will present these tools in order to give an introduction into their conceptualizations, the insights gained – and their respective limits:

- The ReadAct database chronicles real and fictional reading acts and thus allows tracing and analyzing reading practices (and their changes) over time.
- With Shouchaoben, we have created a critical digital edition of a set of unofficial Cultural Revolution handwritten entertainment fiction. Comparing different versions of the same story allows tracing the agency of readers: After all, it was ordinary readers who circulated, copied and re-wrote these stories.
- The DUST database (Database of Ubiquitous Second-Hand Texts) traces the activities of vast amounts of sellers and buyers in the Chinese online second-hand book trade by making machine-readable online posts and diaries in which they communicate about their book-trade related activities.
- Lianhuanhua: Chinese Comics provides the originals and translations of Chinese *lianhuanhua* with critical introductions to each text respectively.
- Reading in China Photographed provides visual sources about actual (and sometimes propagated) reading activities in China, tracing the material conditions and actual social settings in which Chinese individuals read via visual tracking and meta tagging.

Short bio

Lena Henningsen is the PI of the Freiburg based ERC funded project “The Politics of Reading in the People's Republic of China (ReadChina)” and currently a visiting researcher at the China Center, Oxford. She has worked on Chinese popular literature and culture and published widely, including her most recent book *Cultural Revolution Manuscripts: Unofficial Entertainment Fiction from 1970s China* (2021) and translations of a number of Chinese *lianhuanhua* comics (<https://readchina.github.io/comics/>). From 2024 onwards, she will embark on another ERC funded project, “Comics Culture in the People's Republic of China (ChinaComx)” in Heidelberg

3. Fenyé by the Numbers: A Quantitative Analysis of Astrological Contents in Chinese Local Gazetteers

Shih-Pei Chen, Max Planck Institute for the History of Science

Abstract

Field allocation (or *fenye* 分野) is a heaven-earth correspondence system that correlated constellations with discreet geographic regions of the Chinese empire based on the ancient Chinese cosmology of inseparable heaven, earth, and human world. Theories surrounding *fenye* correlations emerged during the eighth to third centuries BCE for uses in political prognostication and continued to develop through



the tenth century CE. As the western astronomical and calendrical knowledge were introduced into late imperial China, it became more obvious that *fenye* did not fit with such modern understandings of astronomical science. However, *fenye* remained ubiquitous in China in the centuries during and after the Jesuit mission. In thousands of surviving local gazetteers—published semi-official geographic overviews of provinces, prefectures, and counties—*fenye* correlations were recorded as a dedicated chapter throughout the sixteenth and nineteenth centuries.

This study uses four thousand digitized local gazetteers and their table of contents to show how gazetteers compilers, usually local literati commissioned by local officials, resisted replacing *fenye* with western sciences. With digital methods, I unwrap this phenomenon in detail and identify the actual rise and fall of *fenye* through a large collection of surviving, digitized local gazetteers. Combining a bird-eye view and close reading, I examine the change of attitudes among local literati during the time when western sciences clashed with the ancient *fenye* theory. I argue that the table of contents of local gazetteers collectively reflects what gazetteer compilers thought this genre was, what function *fenye* served for them and for the local society, and why they persistently kept *fenye*.

Short bio

Shih-Pei Chen 陳詩沛 is a Senior Research Scholar at the Max Planck Institute for the History of Science (MPIWG) in Berlin, Germany. She is a Digital Sinologist with a background in Computer Science. Her research focuses on developing digital research methodologies for historians to work with abundant digital resources, including digital texts, visual materials, and maps. She has been leading the Local Gazetteers Working Group at MPI in the past years, in which her team builds a digital research infrastructure for studying Chinese local gazetteers. The infrastructure, named LoGaRT, uses a framework that treats the genre of local gazetteers collectively as a conceptual database to answer historical questions.

4. Bridging NLP and DH: Beyond metrics to meaning

Baptiste Blouin, postdoctoral researcher

Abstract

The relationship between Natural Language Processing (NLP) and Digital Humanities (DH) is a complex one. On the one hand, NLP can be a powerful tool for DH researchers, helping them to analyze large amounts of text data and extract insights that would be difficult or impossible to obtain manually. On the other hand, there are concerns that NLP models can be opaque and difficult to understand, which can limit their usefulness for DH researchers. This presentation explores this particular relationship, from the NLP point of view, focusing on the issue of model understanding. Arguing that, while performance is important, it is not the only factor to consider when evaluating NLP models for DH applications. Making these models understandable and transparent is also essential, as it enables DH researchers to rely on the results of the analysis and use the models to answer their research questions. Through multiple interactions between historians and NLP specialists, various insights about their interrelationship have been brought to light. It is then necessary to fine-tune the NLP models to the requirements of DH, while at the same time reshaping the expectations of DH itself. These models stand in stark contrast to the resource-intensive algorithms that, while technologically advanced and precise, often lack the comprehensiveness and scope required for tangible DH applications. The presentation examined the contrast between merely using effective models and the inclination towards more interpretable methodologies. Given the undeniable fact that NLP is poised to play an instrumental role in advancing digital humanities, particular attention was dedicated to the solutions proposed within the ENP-China project.



Short bio

Baptiste Blouin is a post-doctoral researcher specializing in Natural Language Processing. His research journey is closely linked to the connection between NLP and the field of Digital Humanities. At the heart of his research is a keen interest in information extraction, with a particular focus on the field of event extraction. After successfully defending his doctoral thesis, Dr. Blouin has chosen to further his academic pursuits with the ENP-China team. This decision stems from his commitment to pushing back the boundaries of knowledge in this interdisciplinary space. His experience and dedication have endowed him with a nuanced understanding of the challenges that arise when trying to bridge the theoretical and practical gaps between NLP and Digital Humanities.

5. Origin, Genres, and Achievements: The Development of Digital History in Mainland China (2000-2023)

Jiang Jie, Shanghai Normal University

Abstract

Short bio

Jiang Jie is currently associate professor at the School of Humanities, Shanghai Normal University. He received his Ph.D. in history from Ecole Normale Supérieure de Lyon. His research field is social history, urban history and the methodologies of digital history. He is co-editor of the journal *Studies of the History of French Concession in Shanghai* 《上海法租界史研究》 (in Chinese, 2016, 2017 and 2019), and executive-editor of the journal *Studies of the Anti-Japanese War in Shanghai* 《上海抗战研究》 (in Chinese, 2019). Recently, he has been working on a project: “How to research history in the digital era” .