

DH BENELUX 2018 – AMSTERDAM, JUNE 6-8, 2018
Integrating Digital Humanities

SHORT PROJECT INTRODUCTION

MarineLives corpus of 17th century legal records.
An example of cross-disciplinary reuse of structured & semi-structured text corpora
Nga Bellis-Phan & Colin Greenstreet

Nga Bellis-Phan (Twitter [@NgaPhB](#))

PhD candidate and teaching assistant in Legal History, Institute of Legal History – University Paris 2 Panthéon-Assas (Paris, France).

Colin Greenstreet (Twitter [@Marinelivesorg](#))

Co-founder and co-director of MarineLives (London, United Kingdom).

General presentation

MarineLives¹ is a collaborative volunteer driven project, founded as a spinoff from a 2012 UK National Archives hackathon. It is dedicated to the collaborative transcription, linkage and enrichment of English High Court of Admiralty's manuscripts, from 1627 to 1677, together with thematically related documents from international manuscript and printed collections.

This corpus of six million words and 12,000 images constitutes a powerful source of commercial, social, material, legal and linguistic insight, and is of relevance to historians, sociologists, linguists, and archaeologists. Our current platform is a semantic media wiki published under a CC BY 3.0 licence, built on our editorial policy of semi-diplomatic transcription, and our folio-based page structure, which relates full text transcription directly to digital images.

Use cases & Potential

MarineLives offers a complete and significant corpus for an extensive scope of time, which enables general conclusions that otherwise could not be drawn from limited corpora constituted by individual researchers.

(1) Legal History

For legal history, our standard edition format contributes to the categorisation of juridical acts by revealing mandatory or optional elements in their structure. The possibility to analyse selected contents can bring to light a historical evolution in legal proceedings or facilitate comparative research on legal practices encountered across European Admiralty jurisdictions. It can also detect influence of the authority on the development of the case, notably with elements deliberately distorted by the same witness's testimonies. Furthermore, a rather complete corpus makes it easier to follow

¹ MarineLives' homepage: <http://www.marinelives.org/>

an entire case. Complex case studies involving parties from different European countries also provide the opportunity to develop a research network of experts from different fields².

(2) Historical & Corpus Linguistics

The corpus presents a precious resource to study the evolution in spelling and lexicography. Thus, a new challenge is to ensure cross-disciplinary studies, enabling corpus and historical linguists to operate as efficiently with the same texts generated³. This requires a technical adaptation of the corpus, either by primarily editing the digitised texts as downloadable XML files, or by allowing ulterior conversion to a different format⁴. The potential for historical linguistics is fruitfully proven with the development of thematic glossaries, which include variations on the same terminology⁵.

Our technology vision: A platform for a virtual archive & research community supporting collaborative scholarship

We are now developing editorial and data standards to maximise the rigour of our corpus and address questions in historical linguistics, legal history, and data science. We are also improving the intuitive interaction of technically novice historians with available search and analysis tools. Our planned collaboration with **Transkribus**⁶, using pattern recognising key word spotting technology, will provide researchers with prioritised manuscript pages relevant to their key word searches, for untranscribed Admiralty Court manuscript folio images.

Our vision for the next five years is to create a virtual manuscript-based archive and associated research community, which will foster a culture of collaborative scholarship. We plan to grow our full text corpus from six million to twenty-five million words and our image collection from 12,000 to 50,000 images. As we scale up our activities, we are actively looking for new archival and library partners, in the United States, London and the Netherlands. To this aim, a charitable incorporated organisation, being registered as **Chronoscopic Education**⁷, will provide a legal, funding and technology home for **MarineLives**.

² A collaborative research team of legal, economic and maritime historians (from the UK, the US, Ireland & France) is working on the case of three silver ships of great economic importance retained by the English authority in the 1650s.

³ Gard B. Jensen, Barbara McGillivray, *Quantitative Historical Linguistics. A corpus framework* (Oxford: Oxford University Press, 2017); Caroline Laske, 'Legal History meets Lexical Semantics: Consideration – the origin of term and concept' (unpublished doctoral thesis, Ghent University, 2017).

⁴ Dr. Sam Kaislaniemi (University of Helsinki), a historian and corpus/historical linguist, is planning postdoctoral research on this theme.

⁵ See our 'Textiles, garments & dyes' glossary in development <http://www.marinelives.org/wiki/Tools: Textiles, garments, %26 dyes glossary>

⁶ Transkribus (<https://transkribus.eu/>) provides a comprehensive platform for the automated recognition, transcription and searching of historical documents.

⁷ Chronoscopic Education (<http://chronoscopic.org/>) will provide the legal, technology and funding home for MarineLives and its sister project on historical mapping, Maphackathon.