DH Benelux

June 7, 2018 – Session 3.2 History II

MarineLives Corpus of 17th century Legal Records.
An example of cross-disciplinary use and reuse of text corpora

Nga Bellis-Phan (University Paris 2 – Institute of Legal History)
Colin Greenstreet (MarineLives)

Presentation text*

This presentation text should be read along with the slides.

Link for SLIDES:

https://github.com/ngaphb/DHBenelux/blob/master/BELLIS%20and%20GREENSTREET%20-%20MarineLives%20Corpus%20-%20Slides2.pdf

[Nga]

[Slide 1] Hi. I'm Nga Bellis-Phan, a legal historian from Paris, France. *[Slide 2]* I joined Colin on the MarineLives adventure as a volunteer transcriber 3 years ago in 2015. It was intended to be a short summer volunteering activity to reinforce my skills in palaeography, and I ended up staying until now because of the brilliant potential of this corpus. MarineLives is essentially a project built on volunteering work of academics but also history enthusiasts. As of now we functions with no money but are registering a charity to be able to receive funding.

General presentation

[Slide 3] MarineLives is a corpus of 17th century legal manuscripts from the English High Court of Admiralty, dealing with maritime cases within a 50 years period. The current achievement is 6 million words of full text transcription and 12,000 digitalized images of manuscripts. *[Slide 4]* The platform adopted by MarineLives is a sematic media wiki.

Each folio of a manuscript has its digitized image and the transcription in full text directly linked on the same page. *[Slide 5]* In order to transcribe and annotate, MarineLives members would go to a standard edition form, that I personally find very user-friendly. In addition, the wiki markup are quite easy to learn. On top are the details of the folio, following by its editorial history. The biggest box is dedicated to the transcription of full-text. Optional fields on the right side allow the transcriber to enter names of people, locations, ships and materials eventually listed on this folio, which is great for data compilation and analysis. For sourcing, each information collected can be easily and precisely linked back to the manuscript's page.

Use cases: Legal History / Historical & Corpus Linguistics

[Slide 6] I am going to share with you two use cases of the MarineLives corpus, first in Legal History and then in Historical & Corpus Linguistics. I hope they will help you get a sense of what we are already able to do with the data offered by the MarineLives corpus but the potential is still huge. You are absolutely welcome to comment on any particular aspect or to propose further steps you think we should take, either for academic research perspective, or technical development goals.

1. Legal History

The MarineLives standard edition format is really useful to legal historians. As you might know, the diplomatic structure of a legal document obeys to an extremely precise order. On this *[Slide 7]* you

can point out the standard structure of the heading in a deposition. You can also distinguish the mandatory information from those only optional. The three zones in colour are mandatory: you have in **blue** the name of the person who made the claim in justice, with a quick description of the claim; in **green** the deposition's date; and in **red** information about the deponent which always begins with a generic sentence saying that he was "Examined upon an allegation given in on behalfe of the sayd claymers", following with the deponent's details (name, address, age, occupation). On the next *[Slide 8]*, you can observe in other manuscripts an identical structure from a diplomatic point of view.

[Slide 9] The important number of manuscripts makes it possible to follow an entire case or a specific person through various cases. After comparing different depositions by the same person, it is sometimes possible to detect information that was distorted or deliberately omitted by the deponent, like in the case of this ship owner named John de Vos. The question here is to know if his ship is from Ostend, which at this time was part of the Spanish Netherlands, or from Amsterdam or cities alike, part of the United Provinces.

[Slide 10] John de Vos was one of the many witnesses in a complex case called the Three Silver Ships case, where these ships transporting goods of great value were retained by the English authority in the middle of the 17th century. The complexity of the case required an international research team, that we rapidly constituted in 2015 and in which I had the chance to participate. We were legal, economic and maritime historians from different countries. This kind of collaborative work really corresponds to the spirit of MarineLives.

2. Historical & Corpus Linguistics

[Slide 11] The second use case would be in Historical and Corpus Linguistics, with the evolution in spelling and lexicography. *[Slide 12]* Historical linguists among you could be interested in our development of technical glossaries, based on authentic extracts from the MarineLives manuscripts. The last achievement was the "Textile, Garments and Dyestuffs Glossary", with up to 1000 terms directly linked to manuscripts quoting them. *[Slide 13]* See this example of the entry "Bombazin".

[Slide 14] Anyway, an important challenge is to make this corpus available to researchers. Not only the data entered in the system should conform to criteria from different disciplines, as each of them are interested in a particular aspect of the manuscript. But, it should be possible for researchers using different technical platforms to proceed with equal ease with the same raw data. Our colleague, Dr. Sam Kaislaniemi from the University of Helsinki, is thinking about creating downloadable XML files from the digitised texts, in addition to the online user interface. This should ensure a possible cross-platform reuse of the same corpus.

[Slide 15] The corpus can also serve for cross-disciplinary common studies. A successful example on Natural Language Processing is the 2014 collaboration between MarineLives and the Informatics Department of the University of Mannheim, resulting in a publication on the study case. They used Named Entity Recognition to isolate the names of persons, locations, and ships, and raised the technical challenges with identifying the ships' name.

[COLIN]

Our vision

[Slide 16-22]

[ADD TEXT]