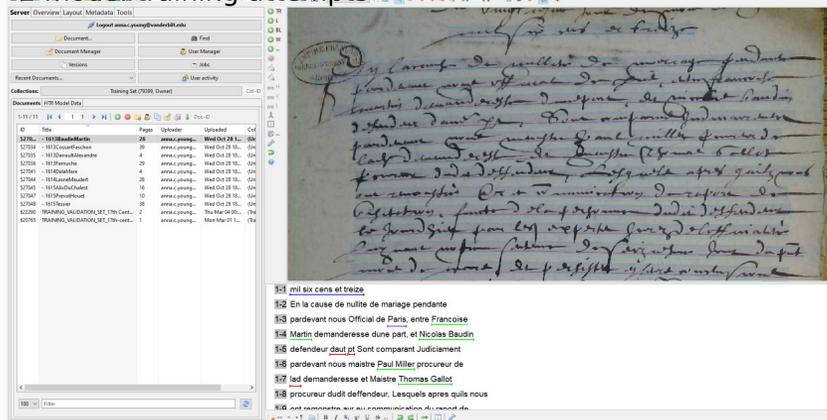


Abstract

This project involved the creation and test use of a Handwritten Textual Recognition (HTR) model based on a corpus of archival documents written in seventeenth-century French. Using Transkribus HTR software, an AI-powered tool used to create machine-readable transcriptions of historical manuscripts, I successfully trained a publicly-available HTR model that may be capable of recognizing a range of similar, seventeenth-century hands from the Parisian ecclesiastical court of the Officialité of Paris. Machine-transcribed documents from the corpus were also made available via a companion Omeka site dedicated to these and other materials from my dissertation research.

Methods

- Software used: <https://readcoop.eu/transkribus/?sc=Transkribus>
- HTR training set: ~23,000 words transcribed manually
- 2 model training attempts

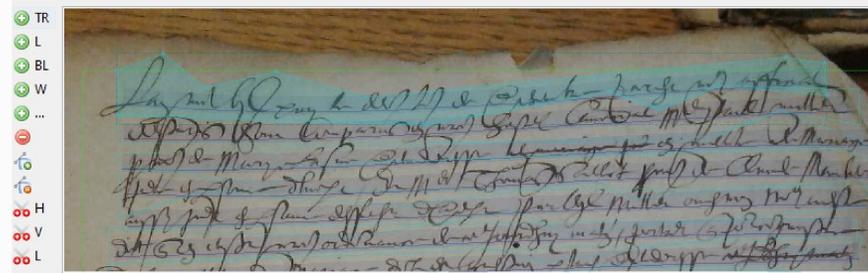


Transkribus logo (upper right) and interface with a sample document (above)

Results

Following two training attempts, the HTR model successfully generated automated transcriptions for a test set of documents of approximately 20,000 words. Compared against manual transcriptions of these documents, these transcriptions were mostly accurate and readable.

However, the estimated Character Error Rate (CER) for the model remains somewhat high, at 20.86%. The model particularly still struggles to accurately transcribe places, names, and strikethrough text, as in the sample transcription below. However, considering the highly repetitive nature of legal documents of this kind, the model was able to capture the basic sentence structure and most used words with few errors. The utility of this model remains to be seen for other similar, seventeenth-century French hands or documents.

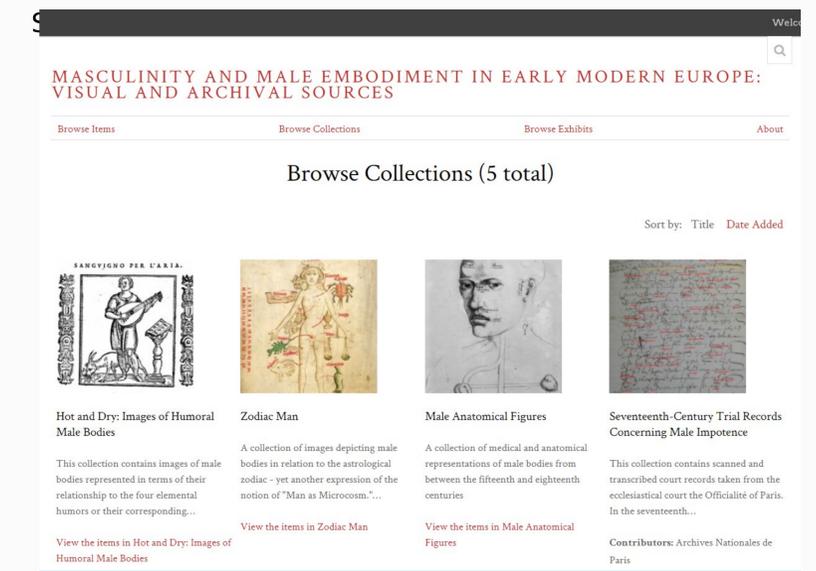


- 1-1 Lan mil vii xii le de Jr de deentre pardnt no official
- 1-2 desaurds sont Comparus en nr histel Cocrat Mer parde nullem
- 1-3 e cr de Marye desne demderesse lemmariage p en nullite de Mariage
- 1-4 prnt en psonne dunept Et Mr Thomas Gallot proc le Cande Naubr
- 1-5 aussy prnt en psonne deffendr daupt par lesd Miller ou p non nous a este
- 1-6 Ait q en enst mr ordonnance de ce Jourdhuy matin portant q po cogoiste
- 1-7 dela nullite du mariage det d Cqeston p lad demanderesse ce sus

Screenshot of sample, machine-generated text, compared with original

Companion Omeka Site

Transcribed PDFs of this corpus are available on the companion Omeka site for this project, "[Masculinity and Male Embodiment in Early Modern Europe: Visual and Archival Sources](#)," a digital exhibition of sources related to the subject of my dissertation: masculinity, normative (and non-normative) presentations of male bodies in early modernity. As the majority of texts included in this corpus discuss legal cases related to male impotence and other male reproductive disorders, the transcribed text of these documents is also included as a resource for other scholars of masculinity, gender, marriage, or social and medical history in early modernity.



Screenshot of Omeka site and collections (above)

Acknowledgements

This work was funded by an Andrew W. Mellon Fellowship for the Digital Humanities

I would also like to thank Sarah Swanz, now Repository Services Manager at Washington University, for her assistance and guidance on this project