

J.R. Ladd  
PhD Student, Washington University  
DHCS Proposal  
27 July 2014

## Named-Entity Recognition and the Collaborative Networks Revealed by Dedications

The upcoming public release of well-tagged EEBO-TCP texts affords scholars new opportunities for the study of early modern language, but the new corpus available for large-scale analysis offers a wealth of social data in addition to linguistic data. Recent early modern network projects, like [Six Degrees of Francis Bacon](#), [Kindred Britain](#), or the now-closed [Making Publics](#), focus on using a combination of primary and secondary material to catalog as many connections between individuals as possible in an attempt to determine the shape of the whole Early Modern social network. The impressive and useful breadth of these kinds of prosopographic projects can be supplemented by the study of more particular, local networks that are discoverable via digital analysis of primary texts. The upcoming EEBO-TCP release makes such network discovery possible.

Using TCP texts made available to me through the Humanities Digital Workshop at WashU, I devised a method for extracting names from the dedications of 17th-century texts using the named entity recognition engine in [Python's Natural Language Toolkit](#). Though printed dedications in the 17th century are not straight-forward repositories of social information, their complex negotiations of identity and identification do provide a window into collaborative networks formed around particular texts. A typical dedication will often include the name of the dedicatee, the name of the author, and then other names that were selected by the dedicator as pertaining to the work in some way: they could be names of the writer's influences, other individuals the writer was trying to impress or curry favor with, or even the author's real or imagined rivals. Whatever the exact character of each relationship, the dedication provides us with a self-selected picture of a particular text's circle, those persons who were deemed important enough to be named in the book's front matter.

Though these self-selected circles cannot give a complete picture of the entire social world at a particular time, they can give us some knowledge about how writers in the period envisioned certain texts and certain persons to be related. Though it may be tempting to look only at the relationships dedications chart between dedicator and dedicatee, these other names also contain useful information about a particular text or relationship. Through these other names, the dedication becomes a site for identifying collaborators, rivals, and influences who might otherwise go without mention. The collaborative, associational forces that constitute any text or set of texts are complex and highly varied, but the dedication is one textual location in which the author or authors begin to sketch out the forces that produced a particular piece of writing.

A paper or poster for the DHCS on this project would include an account of the challenges of using NER on 17th-century texts as well as the presentation of some preliminary results. The first of these challenges is understanding what an investigator is and is not capturing when she only takes those sections of texts given the “dedication” tag. The TCP differentiates dedications from other types of front matter that might be equally dedicatory in nature, such as prefaces, encomia, and epistles. Additionally, variation in spelling and titles makes named entity recognition more difficult to carry out on 16th and 17th century text than it is in the present day. It is perhaps just as interesting for an audience of digital humanists to see where this method fails and in what ways. Finally, there is the issue of what to do with the data once it is captured, and what methods of searching, visualization, and analysis are best suited to the data set.

In many ways this work is indebted to many excellent projects that describe self-selecting social networks in print and manuscript culture, including Ruth and Sebastian Ahnert’s work on [Protestant letter networks](#), Jason Scott-Warren’s work on manuscript networks, and a host of prosopographic projects both digital and analog. What I hope this project can offer or suggest is a new way of discovering social networks from a large volume of TEI-encoded texts.