Renaissance and Reformation
Renaissance et Réforme

Introduction: Computational Approaches to Examining Early Modern Texts

Darren Freebury-Jones

Volume 44, Number 4, Fall 2021

URI: https://id.erudit.org/iderudit/1089352ar
DOI: https://doi.org/10.33137/rr.v44i4.38645

See table of contents

Publisher(s)

Iter Press

ISSN

0034-429X (print)
2293-7374 (digital)

Explore this journal

Cite this document

Freebury-Jones, D. (2021). Introduction: Computational Approaches to Examining Early Modern Texts. *Renaissance and Reformation / Renaissance et Réforme*, *44*(4), 191–195. https://doi.org/10.33137/rr.v44i4.38645

Digital Resource Reviews / Comptes rendus sur les ressources numériques

Introduction: Computational Approaches to Examining Early Modern Texts

DARREN FREEBURY-JONES
The Shakespeare Birthplace Trust

Computational approaches to examining early modern texts have led to discoveries that evaded scholars operating in the pre-electronic age. In the absence of modern electronic corpora, older scholars in authorship or source studies examining phrasal repetitions between plays could never claim that a locution they associated with an authorial candidate or a potential source was unique or commonplace. We can now ascertain just how many times a verbal repetition occurs in texts of the period. We can also determine beyond reasonable doubt whether commonalities between texts exceed expectations, and whether marginal texts fit into the ranges for authorial self-repetition exhibited in attested works, be that the repetition of shared phraseology or vocabulary choices. At its core, this issue concerns digital approaches that seek to uncover minute details of poetic texts and therefore establish whether there is homogeneity between dubious works and the corpora of authorial candidates. But of course, computational approaches to examining early modern texts go far beyond the fields of authorship studies.

Due to the global pandemic, there has been considerably less in-person access to libraries and archives than in previous years, and so digital access has become essential. As this issue reveals, digital resources enable scholars to make new discoveries concerning the lives of important historical figures through examining records of births, marriages, deaths, and wills, and to browse lists and read transcriptions of important seventeenth-century books. Resources that offer opportunities to broaden digital skills have become all the more valuable for not only researchers but also teachers.

As I have alluded, several of the reviews in this issue focus on ways in which researchers can highlight and analyze locutions shared between texts. WCopyfind is a useful anti-plagiarism software that can be employed for this

purpose. The software enables users to highlight strings of words shared between electronic documents. Originally designed to expose student plagiarism, it has the advantage of producing results that are identified objectively, while its results are replicable by any other scholar or research team. It can be set to highlight any specified n-gram (consecutive word sequence) length within a pair of electronic documents, from two adjacent words upwards, and can even identify—within a fraction of a second—approximate matching utterances through adjusting the "Minimum % of Matching Words" and "Most Imperfections to Allow" values. As Dennis McCarthy points out in his review, the software can detect "word-strings that also include a certain small number of unmatched words," but it does occasionally detect some sequences that differ slightly in terms of syntactical arrangement or spelling. McCarthy rightly suggests, however, that users should "normalize many of the most common spelling variations in early modern texts."

The software has been employed by a range of scholars in the field of early modern authorship attribution and source studies, such as Brian Vickers, MacDonald P. Jackson, and McCarthy himself. The option to read matching utterances according to their contexts of use in relation to the surrounding text is particularly helpful for qualitative analysis when it comes to determining whether one document provides a source, or for scholars seeking to investigate the possibility of authorial self-repetition. Here the combined use of WCopyfind with databases such as Literature Online (LION) and Early English Books Online (EEBO) is crucial in ascertaining whether any particular phrasal usage is rare or common in texts of the period.

Brian Vickers has used this plagiarism software in combination with InfoRapid Search & Replace. This "search and find" technology enables users to create a corpus of texts downloaded from databases such as LION and EEBO, and to ascertain whether any utterance highlighted by the software co-occurs with other texts in that corpus. Some researchers have bypassed objective plagiarism software and used the search facilities for LION directly. Mark Hulse reviews the so-called "LION method" pioneered by MacDonald P. Jackson, which involves feeding combinations of words into the LION database in an attempt to establish the authorship of passages or scenes in early modern plays. Hulse stresses the need for "mathematically balanced canons that give prospective authors their fair chance to demonstrate similarities." He calls for more emphasis on control passages extracted from early modern plays and

notes that "Properly performed comprehensive searching can complement traditional qualitative analysis," i.e., reading locutions shared between texts according to their contexts of use. In this way, digital methods can inform more traditional literary-critical analysis.

I have profited much from WCopyfind myself, but I should acknowledge here that, in the case of dramatic corpora, it is impossible to compare results for all plays of the period concurrently and it is a painstaking process to check the rarity of each highlighted phrasal structure by using the search functions in resources like InfoRapid Search & Replace, LION, or EEBO. Fortunately, there have been recent innovations in corpus linguistic studies: largescale corpora such as Pervez Rizvi's Collocations and N-grams, in which all shared phrases in early modern plays are automatically tagged. Similarly, Mel Evans reviews Sketch Engine, an online corpus query system that features several powerful tools for linguistic analysis. For instance, researchers can search for collocations serving different syntactical functions and can compare results across various corpora, even in different languages. One of the most exciting opportunities that Sketch Engine offers early modern researchers is the ability to upload their own corpora. This means that users of this tool can examine "robust, statistically informed and replicable information about the language of early modern texts" as opposed to "[surmising] facts about usage based on erratic search results in EEBO or Literature Online."

Progressing from word n-grams and collocations, Thomas Merriam reviews R Stylo, a suite of programs that facilitates the stylistic exploration of lexical corpora. Merriam proposes that the program stylo() is particularly useful for scholars investigating questions of early modern authorship, and that it excels when it comes to the regulation of linguistic parameters. Stylo() enables users to examine most frequent variables, such as "frequent words (MFW)—single, bigram, occasionally trigram, or the number of 2-character n-grams (MFC), 3-character n-grams (MFC), 4-character n-grams, and so forth," which facilitates an objective analysis of early modern texts. Ros Barber reviews Hugh Craig's version of the Zeta method, which has also been applied to early modern texts as a means of determining authorship. The method attempts to evaluate authorial "style" according to vocabulary choices. Barber stresses that "significant care must be put into experimental design and thoughtful validation procedures" when it comes to such methodologies, especially given the palimpsestic nature of many play texts, which might

contain the hands of multiple authors as well as scribes, compositors, and even actors. Barber demonstrates that results drawn from this method have been frequently misinterpreted in authorship studies, citing such instances as what Pervez Rizvi terms the "bisector fallacy," as well as the possibility that a tested play or portion of a play belongs to a dramatist who is not privy to that test. Barber also notes that mediating factors other than common authorship, such as shared "subject matter, themes, intended audience, literary fashion, a similar education," can account for similarities in vocabulary, and, most crucially, that the Zeta algorithm is genre-sensitive. Barber advises that users of this method should match datasets for size, genre, and period, and points readers towards comprehensive guidance for running and interpreting their own Zeta tests. Marcus Dahl offers some similar caveats in his review of an analytical tool similarly designed to uncover linguistic patterns. DocuScope is essentially a modern dictionary consisting of several million English words. Dahl observes that early modern researchers using this tool will require modernized texts, given the variable spellings in early modern corpora, but he acknowledges that this tool seems especially useful for the field of authorship studies.

Advances in digital technology have led to renewed interest and broadened understanding of the significant Restoration dramatist, Aphra Behn; it is high time that Behn received such careful, scholarly treatment. Joseph Rudman reviews Editing Aphra Behn in the Digital Age, a website designed to make all of Behn's works available to readers, incorporating resources that traverse vast ground including "stylistics, authorship attribution, book history, scholarly editing, chronology, and more." The website is therefore of significant interest to not only readers of Behn's work but researchers in early modern studies as a whole. Yann Ryan reviews the dataset and search interface, George Thomason's Newsbooks, which draws from volumes collected by the seventeenth-century English bookseller George Thomason and includes a browsable list of newsbooks and a downloadable dataset of transcriptions. This search interface will be of interest to scholars seeking to apply digital techniques to early modern sources more generally, and is of particular interest to users who are "keen on the history of the latter of the English Civil Wars, the Commonwealth, and will be of interest to historians of the early English newsbook."

Héloïse Sénéchal reviews the open-access database The Parish of St Saviour, Southwark, which contains all local extant records that relate to the St. Saviour parish between 1550 and 1650. Describing the website as a "digital

Chinese box," Sénéchal points out that the site is "a unique repository of material unavailable beyond the archive, with most documents transcribed and edited for the first time," and observes that "A number of names well-known to scholars of early modern London and of the Bankside theatrical scene are to be found, sometimes revealed in less familiar contexts," including a record of theatre impresario Philip Henslowe as a school governor. This database deserves recognition beyond its application by genealogy researchers, for it can also offer early modern researchers remarkable insights into actors of the period who were living on the South Bank.

Laurie Johnson points out that "'Ancestry.com' might not immediately suggest itself to researchers trained in archival and material historical methods when they are considering an online research tool" but that it is a "powerful tool for historians who might not always be able to access physical collections," which is all the more important in this post-pandemic world. Johnson reveals that his searches of records pertaining to members of the pre-Shakespearean playing company, the Earl of Leicester's Men, have been most fruitful. Ancestry.com has the potential to be a "game changer for this field," and Johnson's innovative research and detailed review helpfully explicate its many uses.

Finally, Rachel White reviews The Programming Historian, a project consisting of lessons in English, Spanish, Portuguese, and French, and comprising a broad range of digital skills from Python to Twitter data analysis. White demonstrates that The Programming Historian enhances digital skills but also "challenges its users to think about the way in which they use them and their research data." The project is of considerable use for research purposes as well as group teaching, providing an excellent resource for groups of any size.

Just as The Programming Historian seeks to enhance the digital skills of users, I hope this special issue will encourage readers to try out some of these resources and will further interactions across the field of early modern studies. The issue consists of contributions from a wide range of researchers, from those at early career stages to established early modern scholars, some of whom have very different research interests and backgrounds. The diversity of contributions is testament to the applicability and usefulness of the resources under review to the field as a whole.

https://doi.org/10.33137/rr.v44i4.38645