Big Data and the Search for Balanced Insight in the Digital Humanities: Macroscopic and Microscopic Reading of Citation Strategies in the Encyclopédie of Diderot (and Jaucourt), 1751-1772

Scott R. St. Louis
Grand Valley State University, stlouiss@mail.gvsu.edu

Follow this and additional works at: http://scholarworks.gvsu.edu/sss

Part of the Digital Humanities Commons

Recommended Citation
St. Louis, Scott R., "Big Data and the Search for Balanced Insight in the Digital Humanities: Macroscopic and Microscopic Reading of Citation Strategies in the Encyclopédie of Diderot (and Jaucourt), 1751-1772" (2016). Student Summer Scholars. 181.
http://scholarworks.gvsu.edu/sss/181

This Open Access is brought to you for free and open access by the Undergraduate Research and Creative Practice at ScholarWorks@GVSU. It has been accepted for inclusion in Student Summer Scholars by an authorized administrator of ScholarWorks@GVSU. For more information, please contact scholarworks@gvsu.edu.
Acknowledgments

This peer-reviewed article was originally published in *Digital Literary Studies*, a scholarly journal hosted by Pennsylvania State University Libraries. An earlier version of the article was presented at the 46th Annual Meeting of the American Society for Eighteenth-Century Studies in Los Angeles, California on March 19, 2015. It was made possible by a generous full-time Student Summer Scholars research grant, provided by the Office of Undergraduate Research and Scholarship at Grand Valley State University.

The author would like to thank Professor Robert Morrissey (University of Chicago) for a letter of support that was included in the grant application, along with Professor David Eick (Grand Valley State University), Associate Dean Gretchen Galbraith (Grand Valley State University), Professor Dena Goodman (University of Michigan), and Professor Emeritus Paul Benhamou (Purdue University), all of whom offered invaluable encouragement and guidance in the completion of the article.
Big Data and the Search for Balanced Insight in the Digital Humanities: Macroscopic and Microscopic Reading of Citation Strategies in the *Encyclopédie* of Diderot (and Jaucourt), 1751-1772

Scott Richard St. Louis (Grand Valley State University)

Consisting of 17 folio text volumes and 11 volumes of engraved illustrations – over 74000 articles and 21 million words – the *Encyclopédie* of 1751-1772 remains a monumental contribution to Western literature (Brewer 449); its publication amid the countervailing pressures of an absolutist monarchy and the Catholic Church has been called “one of the great victories for the human spirit and the printed word” (Darnton 13). Promoting free inquiry into all areas of knowledge and human endeavor, its editors – including Diderot – were threatened with the death penalty for sedition by the French government, and the work was condemned by Pope Clément XIII (Eick 215).1 At least 140 contributors produced this massive corpus (Kafker xv-xxv), and, perhaps due in part to the pressures under which they worked, passages borrowed from other texts are occasionally included in *Encyclopédie* articles without attribution to their true authors or even acknowledgment as quotation. This is a major shortcoming for which the *Encyclopédie* has been criticized since its very inception (Edelstein et al. 213).2 Even so, its accessible framing of philosophical and political ideas (many of which seem as current and crucial as ever) make the *Encyclopédie* a work of enduring interest for cultural historians and literary scholars, some of whom are now utilizing digital technology to develop new insights on the colossal text.3

For example, in April of 2013, scholars Dan Edelstein, Robert Morrissey, and Glenn Roe published a paper in the *Journal of the History of Ideas* entitled “To Quote or Not to Quote: Citation Strategies in the *Encyclopédie*.” The article asserts that the frequent absence of proper attribution in the masterful Enlightenment work often reflects the deliberate use of a shrewd publishing strategy, designed to enable the *Encyclopédie*’s contributors to include in their articles lengthy passages from controversial works unauthorized for publication in Old Regime France (214-215). Calling this a “‘subversive style’ of non-citation” (215), the authors of the article offer compelling evidence – gathered using a formidable array of digital tools – to support their claim that the absence of appropriate citation in the *Encyclopédie* is a phenomenon which was deliberately created at least as frequently as it was caused by a “lack of significant editorial oversight, and a frantic production pace” (220). For example, the authors found that excerpts from John Locke’s *Essay Concerning Human Understanding* were attributed to their author roughly two thirds of the time in the entire text of the *Encyclopédie*, yet they also stated that none of the 38 passages taken from David Mazel’s 1691 translation4 of Locke’s more radical *Second Treatise of Government* included attributive mentions of the work’s title (225).

However, there are at least two *Encyclopédie* articles that *do* include explicit (and positive) mentions of John Locke and his *Second Treatise of Government*: “Démocratie” (Democracy) and “Défense de soi-même” (Self-Defense), both written by Louis de Jaucourt5 and published in the fourth volume of the *Encyclopédie* in October 1754, well before the royal council withdrew the official *privilège* of the *Encyclopédie* in March 1759.6 Jaucourt’s decision to provide readers with such direct mentions of Locke’s *Second Treatise* in these two articles is surprising. After all, Edelstein and his colleagues find that Jaucourt – by far the *Encyclopédie*’s most prolific contributor, having written more than 17000 articles7 – may have employed the subversive non-citation strategy when preparing other *Encyclopédie* articles drawing from the *Second Treatise*, including “Gouvernement” (Government), published with the seventh volume in November 1757 (Edelstein et al. 217-218). The new
evidence I have found therefore suggests that the methods employed by Edelstein et al. should prompt other researchers to build upon and qualify their findings by carrying out the scholarly legwork of *microscopic* or *close* reading, in a manner informed by the *macroscopic* or *distant* reading made possible with the digitized *Encyclopédie* of the ARTFL Project at the University of Chicago. By microscopic or close reading, I am referring to the traditional scholarly practice of examining in depth a passage of text small enough for an individual to interpret without the aid of machines. On the other hand, a researcher who engages in macroscopic or distant reading uses digital tools to process quantities of information too large for traditional methods alone to be of much service. Distant reading therefore includes not only such advanced endeavors as sequence alignment, but also very simple exercises including keyword searches. By saving tremendous amounts of time and energy in comparison to scanning huge quantities of text with the naked eye alone, distant reading can point researchers toward new opportunities for intensive close reading that otherwise might go unnoticed. Macroscopic and microscopic approaches are therefore complementary rather than adversarial in nature (Jockers 171; Wilkens 11).8

This paper argues that the mentions of Locke and his most subversive work in “Démocratie” and “Défense de soi-même” are worthy of serious attention for the insight they can offer to scholars interested both in Jaucourt’s citation patterns and in his working relationship with the chief editor Diderot. While these mentions do not suffice to unsettle the carefully gathered findings of Edelstein and colleagues, they do point to a need for what Edelstein, Morrissey, and Roe themselves call “micro-analysis” (236), which is necessary to (1) differentiate the use of this non-citation strategy from the contributors’ well-attested inattention to the necessary details of attribution, and (2) develop an understanding of how consistently the strategy was used. By way of research that carries out the suggestion made by Edelstein and his colleagues, this paper will demonstrate that the remarkable potential for discovery offered by big data to scholars of history and literature must be balanced by the ongoing practice of a more traditional close reading and erudite sleuthing, which in turn will provide researchers with a more holistic understanding of both the value and limitations of digital tools such as those utilized by Edelstein, Morrissey, and Roe.9 In so doing, this paper will provide *Encyclopédie* scholars with an example of microscopic traditional reading informed by the macroscopic capabilities of a big data tool (the ARTFL *Encyclopédie*) as a method of refining ideas about authorship and citation in the *Encyclopédie.*

To support this argument, it is first necessary to explain the innovative methodology developed by the three scholars, which involved several online databases and a computer program, known as PhiloLine, that is capable of detecting matches between digitized historical texts. The most important resource used by the scholars was the fully digitized version of the *Encyclopédie*, a component of the ARTFL Project hosted by the University of Chicago. Known more formally as the Project for American and French Research on the Treasury of the French Language, ARTFL constitutes North America’s largest collection of digitized French texts (University of Chicago np). Included for free public use in this collection are all 28 *Encyclopédie* volumes, searchable by key word and phrase.

In order to determine whether the digital version of the *Encyclopédie* could be used alongside other tools to detect patterns in the *Encyclopédie*’s plethora of missing attributions, Edelstein and his colleagues began utilizing the PhiloLine program, a sequence aligner which they describe as “an open source data mining extension to the ARTFL Project’s PhiloLogic search engine” (215). Using techniques originally applied in bioinformatics for DNA sequencing, PhiloLine’s algorithms view documents as ordered, user-designed sets of *n*-grams, or groups with an assigned number (*n*) of words taken from a given sequence of text (Edelstein et. al 215-216). By using these sets of *n*-grams (known as “shingles”) in a digital
“reading” of assigned texts, PhiloLine can demonstrate that the same passage has been used in at least two different documents (Edelstein et. al 216).

For their research, Edelstein, Morrissey, and Roe used PhiloLine to compare the text of the *Encyclopédie* with that of roughly 900 French works (all digitized in FRANTEXT, another component of the ARTFL Project) originally published before 1765, meaning that the *Encyclopédie*’s contributors could have accessed them as they were preparing their articles (218). This utilization of PhiloLine, FRANTEXT, and the ARTFL *Encyclopédie* yielded a total of 5,763 results, where each result represented a match between a passage in the *Encyclopédie* and a passage in one of the selected source texts digitized in the FRANTEXT database (218). Edelstein, Morrissey, and Roe then used the sequence aligner on a selection of 1,658 titles contained in the Eighteenth Century Collections Online (ECCO) database, including works written in or translated into French and, again, published before 1765 (218). Finally, the three scholars selected 1,359 French texts originally published between 1527 and 1720 from the “Making of the Modern World” (MOME) database, which provided them with another 4,393 results (219). In sum, by running the sequence aligner on the entire *Encyclopédie* in comparison to thousands of French texts selected from three databases, Edelstein, Morrissey and Roe found more than 10,000 matches between passages in the *Encyclopédie* and passages from French texts which the original *Encyclopédie* contributors almost certainly consulted as they prepared their articles.

Given the immense amount of data that their experiment produced, the three scholars then decided to select the authors and works which they believed would yield the most striking insights regarding the nature of citation (or lack thereof) in the *Encyclopédie*. Thus, they focused on three groups of writers, described in their words as “major Enlightenment authors, including Voltaire and Montesquieu; canonical French authors, from Montaigne to Bossuet; and what might be considered controversial or subversive authors, such as Locke, Hume, and Helvétius” (219). In examining the data which they had collected through this scope, Edelstein and his colleagues effectively confirmed that works which benefited from open authorship and publication authorization by the French royal government were cited far more often in proportion to their overall usage in the *Encyclopédie* than those writings that were published anonymously or were not authorized (224). In other words, passages taken from the works of “canonical” authors such as Bossuet and Montaigne were attributed to their authors far more frequently than passages from more “subversive” writings (Edelstein et al. 223-224). The idea that this pattern is indicative of a clever publishing strategy used by the contributors to sneak controversial material into their *Encyclopédie* articles is further supported by a previously stated fact that I will soon explore in greater detail: Edelstein, Morrissey, and Roe found passages from David Mazel’s 1691 translation of Locke’s *Second Treatise* (entitled *Du gouvernement civil*) used without proper citation in the *Encyclopédie* far more frequently than excerpts taken from the English philosopher’s much less controversial *Essay Concerning Human Understanding* (225).

Simply stated, the experiment carried out by Edelstein et al. has done much to prove that a wily strategy of non-citation existed in the *Encyclopédie* alongside the occasionally careless failure of the contributors to attribute quoted passages to their true authors. Knowing now that the frequent absence of proper attribution in the *Encyclopédie* was likely the result of both deliberate strategy and hurried incaution, scholars must balance the advanced *macroscopic* work of Edelstein, Morrissey, and Roe with simpler forms of distant reading that ultimately facilitate the more traditional *microscopic* work of determining how consistently the strategy of non-citation was used.

Indeed, through even a mere keyword search – a very simple form of distant reading – one can demonstrate the value of using macroscopic tools like the digitized ARTFL *Encyclopédie* to locate potentially rewarding opportunities for microscopic analysis. By
typing “du gouvernement civil” – the name of David Mazel’s 1691 translation of Locke’s Second Treatise – into the search bar on the ARTFL Encyclopédie Project website, one will find thirteen occurrences of this phrase in the entire corpus (the results of this search are available at the following URL: bit.ly/1mlt6Vf). Two of these occurrences are direct references to Locke’s work, and are found respectively in the articles “Défense de soi-même” and “Démocratie,” both written by Jaucourt.

In the first article, one finds this phrase in the last paragraph, embedded in the following quote:

“As for the rights that everyone has to defend their liberty, I am surprised that Grotius and Puffendorf do not speak of them; but Mr. Loke [sic] establishes the justness and extent of this right, in relation to the legitimate defense of oneself, in his work Du gouvernement civil” (Jaucourt, “Défense de soi-même” 816).

In the second article, one again finds the phrase in the last paragraph: “I leave it to readers who wish to expand their horizons still further, to consult … Locke’s Du gouvernement civil” (Jaucourt, “Démocratie” 735). The results produced by Edelstein, Morrissey, and Roe suggest that neither of these articles includes direct quotes from Locke’s Second Treatise: “Not a single one of the 38 [Encyclopédie] passages borrowed from the French translation of Locke’s second Treatise is attributed, or even, for that matter, acknowledged as quotation” (225). However, Jaucourt still mentions this title in both of the articles. If Jaucourt had wholeheartedly committed to a subversive non-citation strategy, then one can reasonably assume that the prolific encyclopedist would not have bothered to mention the controversial Du gouvernement civil in articles where it was not even quoted. Jaucourt’s references to the Mazel translation of Locke in both “Démocratie” and “Défense de soi-même” therefore suggest a need for the microscopic reading that Edelstein and colleagues recommend.

In this instance, close reading illuminates compelling historical evidence suggesting that Jaucourt’s decision even to mention the Mazel translation (albeit while refraining from direct quotation) was a bold one that would not make sense if he had consistently followed a subversive strategy of non-citation. In a 2004 article published in The Historical Journal, S.J. Savonius convincingly argues that David Mazel’s translation of 1691 was likely prepared with the intent of providing its Francophone audience with an anti-absolutist critique of the contemporary regime in France, rather than a mere justification for the revolution that had swept England just two years before (51). Savonius explains that Mazel, a minister at the Protestant church of Gabriac in the Cévennes, had been forced to leave France in order to escape a death sentence passed on him and a number of other Huguenot pastors (58). Mazel fled to Switzerland, but then moved to the Dutch Republic and finally to England (Savonius 58-59). There, it is possible that Mazel collaborated with Locke himself on a French translation of the Second Treatise; evidence is lacking to prove that Locke carefully presided over the production of the translation, but the final page of Locke’s copy of Du gouvernement civil features his handwritten mark of approval (Savonius 68).

The existence of Locke’s mark is particularly striking, given the subversive additions that Mazel made to the text as he translated it. For example, Mazel’s preface to his translation offers an acerbic description of those monarchs and their supporters who would believe that only they can understand how truly to serve God, and who would order soldiers to harm those who do not hold the same beliefs; this aspect of the preface is reminiscent of Mazel’s own experiences with religious persecution in France (Savonius 72-73). The translation appears even more suggestive in light of the fact that the thirteenth paragraph of Locke’s Second Treatise contains no reference to organizing opposition to an absolute monarch, although the corresponding paragraph in Mazel’s work does (Savonius 75; Mazel 14; Locke 230-232).
The evidence gathered by Savonius strongly suggests that Jaucourt’s titular references to the translation at the end of both *Encyclopédie* articles are deviations from a possible strategy designed to prevent the names of subversive works from appearing in the *Encyclopédie*, especially when considered alongside the fact that Mazel’s translation was not authorized by the French government when Jaucourt wrote “Démocratie” and “Défense de soi-même” (Edelstein et al. 224).

Further research must be carried out to determine whether this inconsistency is endemic to Jaucourt’s articles in the *Encyclopédie*. Salient literature suggests that this may be the case, given the chief editor Diderot’s dismissive perception of Jaucourt as something of an intellectual pedestrian, a mere compiler of information rather than an important scholar with unique gifts, admirable passion, and an eminently valuable dedication to the work at hand. Of the prolific Jaucourt, historian Arthur Wilson writes the following:

> “[His] intellect was not creative, but it was retentive, dogged, and quite accurate. His was a truly encyclopedic mind … and while it is easy to scorn such talents, *as Diderot himself was inclined to do*, it ought never to be forgotten that it was … Jaucourt who was as responsible as anyone for making the *Encyclopédie* the great focal point and gathering place of factual information” (Wilson 202, emphasis added).

Given the findings of Wilson, it is plausible that Diderot did not carefully read all of Jaucourt’s articles and thus failed to identify and revise Jaucourt’s deviations from the subversive non-citation strategy in “Démocratie” and “Défense de soi-même.” Bearing this informed speculation in mind, one faces the possibility – increasingly embraced by French Enlightenment scholars – that Jaucourt was not a vapid compiler worthy of Diderot’s contempt, but instead a daring and calculating writer, perhaps with as much courage as Diderot himself. Such a perception of Jaucourt is corroborated by his clear historical standing as an unsung hero of the *Encyclopédie*; he arguably saved the endeavor from failure by authoring over 17000 of its 74000 articles, risking his life by remaining loyal to the enterprise even after the royal council outlawed it in 1759. The thought-provoking findings of Edelstein, Morrissey, and Roe – coupled with a recent augmentation of scholarly interest in Jaucourt, as evidenced by a number of important presentations on his life and work – strongly suggest that now is the right time to search for patterns within (and causes behind) variations of citation and non-citation in Jaucourt’s *Encyclopédie* articles, and perhaps the articles of other contributors soon thereafter.

It is in this way that close reading of the *Encyclopédie* can help to add important dimension to those discoveries made by the use of digital tools. Macroscopic work, including both the advanced experiment conducted by Edelstein et al. as well as simple keyword searches, should prompt scholars to carry out the microscopic work necessary for contextualizing new findings made possible by the use of digital tools. The work carried out by Edelstein, Morrissey, and Roe is groundbreaking for the evidence that it offers in proving the existence of both cleverness and negligence in the *Encyclopédie*’s frequent lack of proper attribution. Most importantly, the caveat that they offer with their results – that “computational approaches to historical texts … must … be tempered by the traditional scholarly practices of ‘close’ reading and intensive analysis of the source material” (236) – is one that should be carefully heeded.

**Works Cited**


Edelstein, Dan, Robert Morrissey, and Glenn Roe. “To Quote or Not to Quote: Citation Strategies in the *Encyclopédie.*” *Journal of the History of Ideas* 74.2 (2013): 213-236. Print.


Notes

1. The introductory paragraph of my paper is similar to that of the third chapter in Eick’s dissertation; these similarities have been included with Eick’s permission.


5 Appropriately, previous scholars have already suggested that the political comments of Jaucourt (by far the most prolific Encyclopedist) “are largely derived from Montesquieu, Locke, and others” and that his contributions to the *Encyclopédie* in general “are filled with acknowledged and unacknowledged excerpts from other sources.” See Frank A. Kafker and Serena L. Kafker, *The Encyclopedists as Individuals: A Biographical Dictionary of the Authors of the Encyclopédie* (1998; republished, Oxford: Voltaire Foundation, 2006), 177.

6 I have included dates related to the publishing history of the *Encyclopédie* with assistance from resources available on the ARTFL *Encyclopédie* Project website, including “General Chronology” (bit.ly/1qehjgC) and “Publication Dates of the Individual Volumes” (bit.ly/V0en8P), accessed respectively on July 16 and August 5, 2014.


9 Edelstein et al., 236: “The macroscopic level from which we were able to discern the use, and often abuse, of authors such as Voltaire and Montesquieu in the compositional process of the *Encyclopédie* speaks to the value of the sort of ‘distant’ reading facilitated by computational approaches to historical texts, a mode of reading that must, nonetheless, be tempered by the traditional scholarly practices of ‘close’ reading and intensive analysis of source material.”

10 Search performed by author on February 17, 2014 via the ARTFL *Encyclopédie* Project.

11 Quote in original French: “Quant aux droits que chacun a de défendre sa liberté, je m’étonne que Grotius & Puffendorf n’en parlent pas; mais M. Loke [sic] établit la justice & l’étendue de ce droit, par rapport à la défense légère de soi-même, dans son ouvrage du gouvernement civil” (emphasis in the original). The translation of this quote included in the main body of this paper is my own.
12 Quote in original French: “Je laisse aux lecteurs qui voudront encore porter leurs vûes plus loin, à consulter le chevalier Temple, dans ses *œuvres posthumes* ; le traité *du gouvernement civil* de Locke, & le *discours sur le gouvernement* par Sidney.” Again, the translation of this quote included in the body of this paper is mine.

13 To quote Savonius: “While it cannot be shown that Locke oversaw the production of the *Du gouvernement*, it can be established that he authenticated the book *ex post facto*. There are two manuscript additions made by Locke to his copy of the *Du gouvernement*: it has written on the title-page ‘Pax ac Libertas’ and drawn on the final page his mark of approval.” Savonius consulted Locke’s copy of *Du gouvernement* in the Bodleian library at Oxford. The observation made by Savonius of Locke’s handwriting on his copy of Mazel’s translation is corroborated by Peter Laslett, ed. *Two Treatises of Government* (Cambridge: Cambridge University Press, 1967), 13 and John Harrison and Peter Laslett, *The Library of John Locke* (Oxford: Clarendon Press, 1971), 40-41 and Plate 6. The plate offers photographic proof of the existence of Locke’s marks on the manuscript.

14 To quote Savonius once more: “[Mazel’s] description chimed with Louis XIV’s billeting of soldiers, mostly dragoons, on Huguenots in southern France rather than with [English King James II’s] violating the law which prohibited the forcible quartering of soldiers in England.”